HumRRO

Final Report

HumRRO
FR-99-23

June 2000

The Transition from KIRIS to CATS, Year 2: Instruction, Communication, and Perceptions at 31 Kentucky Schools

Arthur A. Thacker
Lisa E. Koger
R. Gene Hoffman
Milton E. Koger

Prepared for:
The Kentucky Department of Education
500 Mero Street
Frankfort, KY 40601

Contract Number M-00003669

Human Resources Research Organization
66 Canal Center Plaza, Suite 400 • Alexandria, VA 22314-1591

Acknowledgments

The authors of this report would like to thank the teachers, principals, and district representatives who took time out of their busy schedules to talk with us. Without their cooperation this study would not have been possible. We would also like to recognize the other school personnel and students for graciously welcoming us into their schools and classrooms. The office staff at each of the participating schools handled the lion's share of the coordination for interviews and observations and are to be commended for their excellent work. All of these persons helped us in our attempts to represent Kentucky's schools honestly and accurately within this report. Of course, any remaining shortcomings are the responsibility of the authors. While the authors are appreciative of the comments and suggestions provided by staff members at the Kentucky Department of Education (KDE) regarding the research presented in this report, the opinions expressed herein are those of the authors and do not represent the opinion or policy of KDE.

THE TRANSITION FROM KIRIS TO CATS, YEAR 2: INSTRUCTION, COMMUNICATION, AND PERCEPTIONS AT 31 KENTUCKY SCHOOLS

Table of Contents

List of Tables

THE TRANSITION FROM KIRIS TO CATS, YEAR 2: INSTRUCTION, COMMUNICATION, AND PERCEPTIONS AT 31 KENTUCKY SCHOOLS

Abstract

Thirty-one schools (15 middle, 16 elementary) participated in the second phase of a four-year, four-phase, project examining the transition between the Kentucky Instructional Results Information System (KIRIS) and the Commonwealth Accountability Testing System (CATS). The first phase (Thacker, Koger, Hoffman, & Koger, 1999) included 20 schools (10 middle, 10 elementary), all of which were also included in Phase 2. Schools were selected purposefully to characterize Kentucky geographically and to include a wide range of academic performance levels.

Researchers collected data by observing classes, collecting classroom assessment materials, and interviewing teachers, principals, and district assessment coordinators. The study was qualitative and somewhat exploratory in nature, allowing researchers to investigate related, but unanticipated, topics as they were discovered in each participating school. Interviews were semi-structured (Researchers used the "general interview guide approach" for open-ended questions and semi-structured interviews for demographic information as described by Gall, Borg, & Gall (1996, pp. 309-310).) to allow researchers to adequately respond to each interview situation while also collecting responses on a consistent set of topics. District assessment coordinator interviews were conducted via telephone due to time constraints. Observations focused on the interactions between teachers and students. Analysis of the assessment documents is planned for the end of the study and will focus on the changes in classroom assessment that occur during the first four years of CATS.

The first important finding from this phase is that not much has changed in response to the shift in testing and accountability systems. Schools are continuing their efforts to improve and their methods for doing so are not radically different than they were during KIRIS. Writing continues to be a major focus due to the portfolio component of CATS and the open-response format questions on the Kentucky Core Content Test. The reduction in the number of required portfolio entries seems to have reduced teachers' stress levels, but it has not translated into a proportional reduction in the amount of class time spent working on portfolios. Teachers either reported doing the same number of entries and choosing one less or spending more time editing and polishing each entry. The addition of multiple-choice questions to the calculation of the accountability index has not caused an appreciable reduction in the focus on open-response questions to date. Most teachers reported spending some time teaching multiple-choice strategies and using more multiple-choice questions in class, but concomitant changes in instructional practices were minor.

The consensus of the participants of the study is that the new accountability system is no worse than KIRIS and may be somewhat better. Concerns surrounding CATS are much the same as concerns with KIRIS. Teachers still worry about the consistency and reliability of scoring, student population differences, cohort effects, the breadth of the tested curriculum, test administration procedures, and other issues that might affect either their own school scores or the

scores of schools that they perceive as their competition. Stories of excellent students performing poorly on the test, testing violations at other schools, perceptions regarding the collective intelligence of one class versus another, and other anomalies all feed into a general mistrust and negative impression of the accountability system among participants.

Schools do seem willing to pay heed to information from the Kentucky Department of Education. The Kentucky Core Content for Assessment is clearly driving curriculum. Test results guide schools in forming their Consolidated Plans and in their efforts to improve. Professional development, class schedules, programmatic changes, and resource allotment are all greatly influenced by the Kentucky Core Content Test. This is an interesting contrast—there are complaints about the system and at the same time plenty of hard work to improve instruction and perform better within it.

THE TRANSITION FROM KIRIS TO CATS, YEAR 2: INSTRUCTION, COMMUNICATION, AND PERCEPTIONS AT 31 KENTUCKY SCHOOLS

Introduction

This report marks the second year of a four-year study of the transition between the Kentucky Instructional Results Information System (KIRIS) and the Commonwealth Accountability Testing System (CATS). CATS represents a substantial change in the ways that students are tested and schools are held accountable in Kentucky (Kentucky Department of Education [KDE], 1999a; KDE 1999b), and as such it is important to investigate the consequential impact (as defined by Messick, 1989) of the new testing system. To that end, the Human Resources Research Organization (HumRRO) began conducting visits to Kentucky public schools in the 1998-99 academic year and continued those visits in the current 1999-2000 academic year. The purpose of this research is to investigate the changes schools and teachers are making in response to the changes in the testing system.

The major changes in the testing system, as listed in the District Assessment Coordinator Implementation Guide for the Commonwealth Accountability Testing System (KDE, 1998), include the following.

A. Major Changes and Clarifications for 1999-2000
   - Year of accountability for students participating in the Alternate Portfolio Program[1]
   - Administration Code (a revision of the former Code of Ethics)
   - Possible testing window adjustments (for disasters, calamities, or to observe locally important holidays)
   - Data review/appeals procedures (clarifications only)
   - Kentucky Core Content student listings (describes how individual student test information may be interpreted)
   - National norm-referenced assessment (Students are required to take the norm-referenced portion of CATS and schools are held accountable for it.)
   - Test forms for students with visual or hearing impairments (forms selected to eliminate unfair questions for those students)

B. Major Changes and Clarifications for 1998-99
   - Assessment and accountability terminology (updated and clarified)
   - Kentucky Assessment Service Center (clarified hours of operation and telephone number to call for help or with questions)
   - High school testing (changed grades that certain portions of the Kentucky Core Content Test are given)
   - Two-week testing window (for Kentucky Core Content Test and national norm-referenced assessment)

[1] The Alternate Portfolio Program was designed as an assessment system for students with severe disabilities. It is a compilation of their academic work and is only an option for those students without access to the Kentucky Core Content Test.

- Possible adjustment of the testing window (for natural disasters or other calamities)
- Performance level increments and credit (Novice and Apprentice performance categories are split into 3 sub-categories each and points are awarded toward the schools accountability index accordingly.)
- Computing the accountability index (weightings for content area tests and non-academic components were changed and now include multiple-choice items)
- Order of testing (Test administration order may not be changed.)
- Calculator issues (clarified, school's responsibility to provide calculators to students)
- Students with disabilities (clarification only)
- Use of extraneous/additional materials during assessment (clarification only)
- Use of incentives during assessment (clarification only)
- Student related disciplinary matters (clarification only)
- One page limitation to open-response item (previously unlimited)
- Writing portfolios (reduced the number of pieces of evidence required per portfolio)
- Alternate portfolios (also fewer required pieces of evidence)
- Core Content Report (expanded results-added content scores for arts and humanities and practical living/vocational studies)

It would not be surprising, given the magnitude of some of these changes, that schools and/or teachers would alter their practices. Originally, the testing and accountability system was designed, in part, to stimulate improvements in instruction. Changes in instructional practices resulting from the implementation of the accountability system are well documented (Harris, Hoffman, Koger, &Thacker, 1998; Hoffman, Harris, Koger, & Thacker, 1997; Kelley & Protsik, 1996; Koretz, Barron, Mitchell, & Stetcher, 1996; Stecher & Barron, 1999).

Methodology

Case Study Description

This research is best defined as a case study. It is not typical of case study research due to the implicit need to generalize from the schools that were chosen to the larger system. Generalization like this is not typical of case study research but applies for this particular study if we examine the caveats and limits of the study. First, the case that was studied was the Commonwealth Accountability Testing System (CATS), and not any of the schools that we visited. The schools represent the primary level of actors that were observed to learn about the case selected. The secondary levels of actors were the classes within the schools and the tertiary actors were the many teachers, administrators, and district personnel we interviewed. In order to learn about the system, we studied the schools, the classes, and the people who comprise them. This study can be compared to a more typical education case study where researchers study classes, principals, teachers, and students in order to learn about a school. Our study simply started one level higher and examined the system itself (Thacker, Koger, Hoffman, & Koger, 1999, pp. 8-9).

Second, the study was evaluative of the case, but not in traditional case-study terms. Instead, this study examined an aspect of consequential validity, as defined by Messick (1989). In

short, this study sought to discover the impact CATS had on the everyday practices of teaching and learning in public schools. Particularly, this study sought to determine the impact of changing the accountability system from KIRIS to CATS on classroom practices.

This idea leads to the necessity for generalizations to be made from the results of this study. The common problem with generalizing from case study research is that often one case does not act like another. In this instance another case would be another state's accountability system, and that is not the level of generalization with which we are concerned. In case study research it is common to find that actors within a case will react to external factors in similar ways (Stake, 1995). For example, students will tend to exhibit certain behavior patterns within a specific school environment. We might therefore expect schools to exhibit certain patterns within the CATS system. Generalization from these actors within CATS may tell us a great deal about the expectations and reactions of the larger system.

Sampling

Researchers visited 31 Kentucky schools representing 15 districts. Each of the Regional Educational Service Areas defined within the state was represented in this study. One middle school and at least one elementary school were visited from each participating district. Elementary grades from one of the participating districts were split between an elementary school, housing grades 2-4, and an intermediate school, housing grades 5 and 6. Twenty of the 31 selected schools participated in Phase 1 (Thacker et al., 1999) of this project and were also visited during the 1998-99 academic year.

The first factor taken into account in choosing cases was the location of the school within the state. There is a great deal of diversity among public schools within the Commonwealth of Kentucky. It contains rural, urban, and suburban schools. It contains independent and county schools. A section of Kentucky is located in Appalachia, which has been identified in other research as an academically "at-risk" area of the country (Baldwin, 1994; Branscome, 1970; Wimberley & Morris, 1996). The sample was chosen to include schools that characterize the geography and diversity of the state. Efforts were also made to ensure that no region was represented by a single portion of the academic range of schools in the sample. For instance, not all low-performing schools in the sample are located in the Appalachian area.

Schools were then selected using a stratified purposeful sampling technique (Patton, 1990; Gall, Borg, & Gall, 1996) for Phase 1 of the project. For Phase 2, schools were added to the original list using the same process. Exclusion criteria were used to limit the possible choices before selection began. Those exclusion criteria include:

1. Class population. Due to logistics and economic concerns, schools with a class population of less than 50 students were excluded from the study. This stipulation facilitated timely data collection by placing researchers only in schools with several teachers from whom to collect data. This stipulation also ensured that each school visited was represented by multiple data sources, which helped assure the validity of the findings of the research.
2. Separate Middle and Elementary Schools. A portion of the original KERA legislation

decentralized many of the school operational decisions. Schools, with their SBDM councils, now make many decisions independently. Those decisions were, to a large extent, the focus of this research. If a school shares a building, a principal, and perhaps even an SBDM council with another school—as is the case in many combined schools—it becomes difficult to distinguish one school from the other. Since CATS stipulates schools as the primary level of accountability and tests each subject at elementary, middle, and high school levels, the sample was limited to schools clearly distinguishable as elementary or middle schools.

3. Willingness to participate. Participation in this study was not mandatory. There was an element of self-selection in the sampling technique. If a school or district chose not to participate it was replaced with a similar school or district from the same Regional Education Service Area. The results of the study were influenced by the exclusion of those schools to an unknown extent. Similarly, if a specific teacher or teachers within a school chose not to participate in the study, the degree to which their exclusion influenced the conclusions of this research is unknown. Fortunately, very few schools/teachers chose not to participate in the study. Only one district chose not to participate in Phase 1. One selected school was also replaced in Phase 2.

In addition to the exclusion criteria, other factors were used in the selection of schools and districts. The sample was chosen to represent an appropriate range of school assessment scores. Since the original 20 schools were chosen before the Kentucky Core Content Test, KIRIS scores were used for this purpose. The 11 schools that were added for this phase were also selected using 1998-99 KIRIS scores as a criterion. The selection process occurred before the Kentucky Core Content Test scores were released. Once 1998-1999 scores were released, they were examined in relation to the state mean in order to ensure that the sample chosen was not skewed toward the high or low end of the range of Kentucky schools. A comparison of sample and state means, as well as descriptive statistics for the sample, is provided in Table 1. Thirteen of the 31 schools chosen had index scores above the state mean.

Table 1. Overall Index Scores for Schools

1998-1999 Kentucky Core Content Test	Elementary Schools	Middle Schools
Kentucky Mean	61.3	54.3
Sample Mean	65.2	53.1

The sample mean index score for elementary schools was higher than the state mean. We had less freedom to select elementary schools based on their scores due to the necessity of selecting schools with at least 50 students per grade level and an attempt to choose those elementary schools that fed students into the selected middle school. Once confirmation of a middle school was reached, the criterion for choosing elementary schools limited our options. The range of elementary school scores is reasonable and, in fact, eight of the elementary schools selected scored below the state average and eight scored above. Three very high-scoring elementary schools help account for the higher sample average.

The last criterion for selecting schools for the sample only affected those schools replacing schools that chose not to participate in the study. The Kentucky Department of Education, in

partnership with CTB/McGraw Hill, is conducting another research project in cooperation with HumRRO. That study has the potential to be informed by the results of this one. It involves retesting students from a sample of schools and is commonly referred to as the "growth study" by KDE and CTB/McGraw Hill (report forthcoming). The sample of schools chosen for their study is much larger than our 31 schools, but there was an effort to make sure that their sample was inclusive of ours. HumRRO provided KDE with the projected sample for this study. The sample drawn by KDE for their study included those schools. When one of those schools declined to participate, HumRRO selected a replacement school from the KDE sample wherever possible; however, in one instance there was no suitable replacement in the KDE sample.

This study focused on science and social studies teachers for data collection. Teachers specializing in other subjects were not included unless they also taught science or social studies. Science and social studies were studied in particular for the following reasons:

1. Science scores at the middle school level have not enjoyed the same degree of growth as other subjects. This makes science instruction at the middle school level of particular interest.
2. Reading and mathematics have been the subject of other studies of the Kentucky education system (Borko & Elliot, 1999; Stecher, Barron, Kaganoff, & Goodwin, 1998; Wolf & McIver, 1998). HumRRO included mathematics teachers in some of its earlier studies of the Kentucky system as well (Harris et al., 1998; Hoffman et al., 1997).
3. The Kentucky accountability system gives social studies about the same emphasis as mathematics, reading and science. Arts and Humanities and Practical Living/Vocational Studies are tested using fewer test questions and they account for a smaller portion of each school's accountability index (KDE, 1999b). Schools also might not have teachers specifically designated to teach those subjects.
4. Science and social studies are tested in different academic years at the middle and elementary levels. Science is tested in the fourth and seventh grades. Social studies is tested in the fifth and eighth grades. The ways that accountability and testing function at different grade levels, both in terms of what the scores mean for the school as a whole and what those scores mean for individual students, have the potential to be very different.
5. Only studying two content areas limits the number of teachers in the sample from each school. By attempting to study every teacher at a school who teaches science or social studies we avoid the problems associated with requesting volunteer teachers or identifying randomly selected teachers prior to scheduling and conducting the school visit.
6. The HumRRO researchers involved in this study have more experience studying science and social studies than other content areas in Kentucky. Our familiarity with the content and the literature surrounding these subjects helps HumRRO researchers establish rapport with the respondents, gain their trust, and understand respondents' language and culture.

Accessing the Data

Selected schools were initially contacted by mail. They were sent a letter of introduction

inviting them to participate in the study (Appendix A) and a copy of the summaries of the results from Phase 1 of this project (Appendix B). Schools participating in Phase 1 of the project and newly selected schools received the same letter. The letters were followed by phone calls to the principals of the selected schools. Researchers verified school participation and scheduled visits during this conversation. School principals were given an opportunity to ask questions regarding participation during this conversation. They were also provided a toll-free telephone number that they could use for asking further questions or addressing concerns that occurred after the initial contact.

Once schools agreed to participate in the study they were sent a packet of materials designed to make the visits as unobtrusive and productive as possible. The packet included a letter of introduction for participating teachers (Appendix C), a schedule form designed to facilitate interviews and classroom observations (Appendix D), and a request for logistical information such as area hotels, parking regulations, school hours of operation, etc (Appendix E). Introduction letters for teachers also contained contact information so that teachers could ask any questions and address any concerns they had prior to the research team visiting their school. The introduction letter also reminded teachers that researchers would collect assessment documents from them during the visit.

A member of the school's administrative staff typically created the initial schedule of interviews and observations. Researchers adjusted the schedule to account for staff absences, lengthy interviews, and other unforeseen events at the school. Interviews were given priority over observations when time constraints forced researchers to choose between the two (per an agreement with KDE). Group interviews were discouraged; however, they were necessary in a few isolated instances. When departments or teams of teachers had common planning periods, they were necessarily interviewed together. Group interviews only occurred in 6 of the 31 schools. No more than 5 teachers were interviewed together in any instance and that only occurred one time. Two groups of 4 were interviewed. All other group interviews were groups of 2 or 3. Administrators were not present during teacher interviews.

Data Collection Methods

Three types of data collection were chosen for this study. The primary method of collecting data was through direct interviews with teachers, principals, and district personnel (District Assessment Coordinators). All interviews were conducted using a general interview guide approach (Patton, 1990). Briefly, the method involves outlining a set of topics to be explored with each respondent. The order in which the topics are explored is not necessarily predetermined and interviewers can alter the exact wording of the questions to facilitate clarity and limit ambiguity. They might also restate the question or follow up with other questions dependent on each interview situation. This type of interview approach is very appropriate given the somewhat exploratory nature of this research project.

The topics for the interviews were written in question format (Appendixes F, G, and H). They were drafted based on the results of Phase 1 of this project and edited and clarified through a peer review process. All reviewer comments and suggestions were considered and either used to alter the interview instruments or as points of clarification among the interviewers prior to the start of the project. Interviewers met regularly throughout the project in order to ensure the consistency of the content of the interviews. District-level interviews were conducted via telephone due to time constraints.

The second type of data was collected during classroom observations. Researchers observed selected classes (necessitated by schedule constraints, willingness to participate, and other factors) taught by a portion of the interviewed teachers (Observations were not possible at all in one of the participating middle schools.). They used a continuous recording methodology (Gall et al., 1996) in keeping observation notes. They focused primarily on descriptive observational variables (where the teacher was located in the classroom, references posted on the walls, the topic studied in the class); however, they were encouraged to note inferential observational variables (whether the students were attentive, the taxonomic level of questions posed to students, descriptions of group interactions among students) as well. Whenever possible two researchers would observe the same class in order to check for observer agreement. Observer effects were limited by using four researchers in two-person teams and reconstituting the teams regularly. No school or district was represented by only one researcher. Researchers regularly discussed descriptive variables with regard to the inferences and interpretations that stemmed from these data.

The final type of data collected was documentation. Interviewed teachers were asked to provide three assessments that they used in their classrooms. They were asked to provide the lowest level of assessment they used (quiz, daily journal assignment, etc.), a middle level assessment (chapter test, unit test, project requirement, etc.), and the highest level of assessment they used (final examination, mid-term examination, large project requirement, etc.). These documents were cataloged and stored for content analysis at the end of this project.

Assessments were chosen over other documents because of their availability and their direct use in the classroom. Analyzing assessment documents also helps researchers meet the

stipulations of effective qualitative document analysis. In order to fully understand the document the researcher must study the context in which it was produced. That context includes the author's purpose in writing or using the document, the author's working conditions, the author's intended and actual audience, and the audience's purpose in reading the document (Gall et al., 1996). These conditions are readily explored given the nature of classroom assessment documents and the other data collected during the course of this project.

Data Analysis

Interviews were analyzed (category analysis, Denzin & Lincoln, 1998) using a five-step process. The first step involved a single researcher reading all collected interview data and developing a set of categories that characterized the responses to an interview question. That researcher then used those categories to describe a portion of the interview data. Second, another researcher read a portion of the interview data and revised and expanded the categories to better capture the data. Third, all four researchers used the categories from the second step to analyze a portion of the interview data. They used a consensus-building technique to revise, clarify, and expand the analysis categories. Fourth, the four researchers analyzed a larger portion of the interview data noting exceptions to the categories and additional required editing of the category descriptions. Agreement between researchers was examined and consensus building was used to establish parameters that could be readily agreed upon with regard to the data collected and the categories produced. The last step was to analyze the full data set. This was accomplished by one researcher and then independently checked by the three others in sections. The entire process was repeated for each interview topic.

Observation notes were analyzed using a reflective analysis procedure (Gall et al., 1996). This procedure relied heavily on the judgment of the observer regarding first what details to note from the classroom observation and second how to interpret those details. Once observation notes were taken and observers had written reflections on the meanings of those notes, the notes and the classroom observations were discussed among the other researchers, particularly the researcher who visited the same school (and/or classroom). Possible misinterpretations of the observations were discussed and debated. The observation data was compared to interview data from the observed teacher and from other teachers and the school principal. The written reflections of the researchers were appended based on these considerations. The initial interpretation of the observer was not rewritten; however, any use of the observation for illustrative purposes in the narrative of the results section of this report utilizes the observer interpretation as well as appended information. All interpretations from observations were examined with regard to other data for purposes of triangulation and confirmation that the interpretation was reasonable. Similarities among observed teachers were noted. Exceptions to these similarities were also noted. Both common and rare practices were used for the construction of the Results section of this report.

It should also be noted that the report itself is not constructed specifically by interview topics. Tangential information was often provided by participants and frequently led to pertinent discussions regarding the actualization of the testing and accountability system at the school or classroom level. Many participants came to the interview with their own questions and concerns

about the Kentucky Core Content Test. Therefore, we allowed our preconceived structure to be reorganized by the structures of our educators, which in turn allowed us to address the impact of CATS as seen by those being impacted.

The answers to interview questions were not always consistent from district to district, school to school, or even within a single school. In fact, not all participants answered all questions. If participants chose not to answer, the interviewers did not press. In some instances, the interviews were cut short because of teachers' other responsibilities. In other instances, interviews were conducted under difficult conditions (during a class while students were completing an assessment, on the playground while the teacher supervised a class). In one school, the time that was allotted per teacher interview was only 15 minutes. The interviews in that school were shortened considerably. Participants' concerns varied greatly. This report does not attempt to quantify the numbers of teachers mentioning any single issue, but instead describes the issues as participants discussed them. Often this report may seem to contradict itself, for instance, reporting first an area of concern among teachers and later referring to that same issue as a perceived strength of the system. This is not a contradiction in the report, but rather a counterpoint made by one or more participants. It is very possible for the same aspect of the accountability system to be perceived as both a strength and a weakness depending on the vantage point of the participant. Rather than force a quantitative statistical methodology on the data in order to determine the prevalent perception, we chose instead to provide both views, and to treat each as relevant and genuine.

Results

Demographics of Schools

Table 2 lists pseudonyms for schools that participated in both of the first two phases of this research. It also includes a short description of each school that indicates whether the school is rural, suburban, or urban, where the school is located in Kentucky, and a very short identifying comment. This table makes no attempt to fully describe the schools. It is provided as a reference due to the relatively large number of schools mentioned later in the report. A more complete description of the schools can be found in the text of the Phase 1 report (Thacker et al., 1999).

Table 2. Participant School Descriptors (Schools Visited Last Year)

District	*School*	*Descriptors*
Elm	Elementary	Large, new, rural, Eastern
	Middle	Large, competitive, rural, Eastern
Oak Independent	Elementary	Suburban, writing emphasis, Central
	Middle	City school with some urban aspects, 7th and 8th grade only, Central
Pine	Elementary	Urban, diversity requirements (busing), Central
	Middle	Urban, poor academic reputation, high percentage of special education students, Central.
Spruce	Elementary	Rural, returning to more traditional teaching after reform emphasis, Western
	Middle	Rural, technology emphasis, Western
Cedar	Elementary	Suburban, older school, experienced staff, Northern
	Middle	Suburban, extreme teaming strategy and schedule, Northern
Hickory	Elementary	Rural, small town school amid several other "county" schools, Central.
	Middle	Rural, small town school, Central.
Poplar	Elementary	Rural, diverse population, switching 3rd and 4th grade teachers, Western.
	Middle	Rural, dropping student population, Western.
Cottonwood	Elementary	Rural mountains, mining community, Eastern.
	Middle	Rural mountains, double classes in assessment grade subjects, Eastern.
Locust	Elementary	Urban, innovative, older building with large classrooms, Central.
	Middle	Urban, contesting KIRIS scores, Central.
Walnut	Elementary	Rural, K-6, returning to basics, Southern
	Middle	Rural, 7-8, Southern.

Table 3 lists pseudonyms for schools that were added for the second phase of this research. These schools did not participate in Phase 1 of the research. The information provided in Table 3 is the same as that found in Table 2. A more complete description of these newly added schools is found immediately following the table. These two tables demonstrate the variety of schools contained within the sample group.

Table 3. Participant School Descriptors (Newly Added Schools)

District	*School*	*Descriptors*
Maple	Elementary	Rural, Strong teacher/principal cooperation, Central
	Middle	Rural, emphasizing multiple intelligences, Central
Chestnut	Elementary	Urban, routine team teaching, Central
	Middle	Urban, writing emphasis, asymmetric schedule blocks, Central
Willow	Elementary	Suburban, strong academics, low SES, Western
	Middle	Suburban, just switched to block schedule, Western
Dogwood	Elementary	Rural, large number of impoverished students, South Eastern
	Middle	Rural, teaming in pods, South Eastern
Cypress	Elementary	Rural, grades 2-4, student work prominently displayed, South Central
	Intermediate	Rural, grades 5-6, mathematics emphasis, South Central
	Middle	Rural, grades 7-8, split between 2 buildings, South Central

Maple County School District

Maple County is located in North Central Kentucky. It contains three elementary schools that feed into one large county middle school. The district's role in helping schools prepare for the Kentucky Core Content Test in Maple County included scrimmage tests that were composed of released items from the previous test coupled with a review of student performance (They also purchased practice materials from CTB.). In this district the elimination of the common released questions was viewed as a detriment to the system.

Maple County Elementary School

Maple Elementary is an older building, but pleasant on the inside. Students' work and other displays were common. The school staff was generally composed of veteran teachers with a high energy level.

Maple's principal was a former mathematics resource teacher. Her focus at the school was on integrating the curriculum and slow steady progress. She explained, "We need more analysis of test scores coupled with analysis of current students' work. We can't knee-jerk to released test

questions. We need to look at the Core Content for Assessment." That sentiment, as well as her instructional leadership, helped create an atmosphere of mutual respect between teachers and administration at Maple. More than one teacher referred to the interaction between the teachers and principal at the school as "synergistic."

The observed classes at the school combined both traditional and reform-oriented practices. There were group activities and discussions along with spelling words and essays. Students were assigned a wide variety of tasks at Maple. Teachers expressed their high expectations for student work and expected homework to be completed and turned in on time. Teachers often worked in two-person teams designed to facilitate cross-curricular projects and teaching.

Maple County Middle School

Maple County Middle School is located on the same campus as the county high school, although it is in a separate building. Each grade is located in a separate part of the building. The school underwent a major renovation five years ago when the sixth grade was moved there from the elementary schools.

The school's principal and vice-principal were in their second year. The current principal served as vice-principal at Maple before. He was very positive about his school and public education generally. Maple has been assigned a Highly Skilled Educator (HSE), although that was rarely mentioned during interviews.

The academic focus of the school is on Howard Gardner's (1983) Multiple Intelligences[2]. The principal and several teachers described the difficulty meeting the needs of the students. The principal described students who could take apart, repair, and rebuild a small engine easily, but who struggled in math. The strategy at the school seems to be to teach toward these "less academic" talents to motivate students to strive toward a more typical curriculum.

The principal explained that the district provided a "Freshman Academy" to help students transition from middle to high school. He would like to see the process help those students transitioning from elementary to middle school as well. He also commented on a disparity in the test for middle school as opposed to other levels. "The average index for middle schools is lower than for elementary and high schools. This leads to a poor public perception of middle schools." The principal also questioned the developmental appropriateness of the test and lamented about the effect that testing has had on the public perception of the education profession.

Chestnut County School District

Chestnut County is in the central portion of the state. It contains a small city with an

[2] Gardner's (1983) work suggests that intelligence is not a single parameter that can be identified with a number on an I.Q. scale. He provides evidence that there are at least 8 separate intelligences and that children vary among those intelligences. He further suggests that different children will learn at different rates dependent on the modality of the instruction, or the particular intelligences the instruction seeks to access.

independent school system. County schools compete with independent city schools for students. The student populations of the schools we visited in this county were reasonably diverse in terms of ethnicity. Teachers reported that there was a wide range of academic and socioeconomic diversity as well.

The district office in Chestnut County assists schools with curriculum alignment, which has a large emphasis in the county, and helps schools adopt textbooks that are closely aligned with Kentucky's Core Content for Assessment. Curriculum alignment efforts in the county include unit planning as well as yearly curriculum planning. Professional development in the county was strongly tied to the assessment. On-demand and open-response training is common. The district assessment coordinator maintains a list of commonly used testing vocabulary that is provided for all schools. Scrimmage testing was done in all schools in the county. Ethics training has also received a good deal of emphasis.

Testing special education students was a concern in this district. Specifically mentioned were difficulties managing the alternate portfolios, logistical issues involved in getting all the testing materials out and back while maintaining security, and the appropriateness of the students' Individual Education Plans (IEPs). Ensuring the appropriateness of the IEPs included both matching the IEP with the correct type of accommodation and making sure that the IEP matches the severity of the disability and has not been written just to provide a testing advantage.

Chestnut County Elementary School

Chestnut Elementary is a county school located within a city that also contains an independent district. The school building has an "open-school" design but partitions have been erected to give it more of a traditional enclosed classroom atmosphere. One teacher joked that this was the first of her 15 years at the school in which she had a classroom with walls. She was very positive about the change, claiming that it provided fewer distractions for her students. Part of the school was remodeled this year and many of the large open areas were broken into classrooms with permanently constructed walls. An open library/media center was located in the center of the building. The administrative area was also in an open area near the front of the building.

The school has recently constructed a walking trail with the help of many of the students' parents. It is used for exercise and as a "nature walk." The halls in the school were painted with murals and timelines geared to Core Content.

The population of the school was ethnically and socioeconomically diverse. Observed teachers seemed to set high standards (as evidenced by the difficulty and volume of their students work compared to other schools in this study) and strive to help students attain those standards. The school has chosen Great Books[3] rather than Accelerated Reader[4] as its dedicated reading

[3] Great Books refers to a reading program produced by the Great Books Foundation (Information available at www.greatbooks.org), a non-profit education organization. The program promotes reading and discussing a selection of books and offers teacher training for conducting "shared inquiry" lessons.

[4] Advantage Learning Systems, Inc. (www.advlearn.com) designed the Accelerated Reader Program. It is used in many Kentucky public schools to promote reading and reading comprehension. Students read prescribed books

program. One teacher explained that, "Great Books is typically reserved for Gifted and Talented classes, but at this school, everyone uses it."

Chestnut County Middle School

Chestnut Middle is a large school, housing about 750 students. The building itself is older and in need of repairs in several places. Each of the three grades taught at the school has its own schedule. No grade uses a standard 6 or 7-class period a day schedule. Instead, all have some sort of block design. Also, at each grade, the block of time devoted to language arts is longer than the other core subject blocks.

One of the most interesting aspects of our interview with the principal at this school was a statement regarding his perception of the role of the school. "We need to get to the point that it's not the population that counts most (in determining test scores), but the instruction. That's the hard part because there are no more resources for needy students." The idea that instruction and resources can overcome differences in socioeconomic status is not new, but it is rarely stated in so clear a manner at public schools. The other major implication of the statement is that the principal believes that there are instructional practices, which are different from the norm at his school, which should be promoted for this purpose. When asked which practices needed to be emphasized in order for this to occur, the principal mentioned higher order thinking skills and matching instruction to the learning modality of the students. Howard Gardner's (1983) work on multiple intelligences seems to have influenced the principal's goals to a great extent. He mentioned multiple intelligences and learning modality frequently during his interview.

The school's emphasis on open-response format questions seemed fairly new this year. The front office began collecting open-response questions from teachers as well as examples of student work in order to verify that they are being used in all classrooms this year. Several teachers expressed concerns about the school's strong emphasis on writing. The language arts classes at the school have about one third fewer students than other classes and last about one fourth longer. The rationale for creating such a disparity was the need for individual student conferencing during the preparation of the portfolio. The principal, however, clarified the heavy emphasis on writing on the state assessment. "If they can't write, we know they won't score well. Writing skills should help all areas on the test," he explained.

Willow County District

Willow County is a Western Kentucky county containing one small city. Several of the state's top-performing elementary schools are located within this county. The schools do not all serve a high socioeconomic population, and the county schools have recently been the subjects of other studies regarding their methods. In fact, it was difficult to schedule one of these schools for this study because of the number of activities in which the schools had already agreed to participate.

---

and take quizzes or tests to determine their comprehension levels. Students are awarded points, that are often tied to a reward system, based on the length and difficulty of the books.

Willow Elementary School

Willow Elementary School is an eight-year-old county school located on the outskirts of town. Because of its high number of students who receive free or reduced price lunches, it qualifies as a schoolwide Title I program. The principal is in her second year; she is the school's fourth principal.

Despite the high percentage of free/reduced price lunch students, this school made a strong showing on the 1999 Kentucky Core Content Test. Despite their strong scores, particularly in writing, the principal stated that their science scores had declined somewhat and that reading had not improved as much as they had been expecting.

The school has grouped its classes in a K-1, 2-3, and 3-4 arrangement. Fifth-grade classes contain only fifth-grade students. This year's blended 3-4 grade classes marked the return to this arrangement after experimenting last year with separate third and fourth grades. Teachers blamed the decline in science scores on the separate third- and fourth-grade classes, stating that under the blended system more students would be exposed to two years of fourth-grade science. With the addition this year of a third teacher at the fifth-grade level, the fifth grade is more departmentalized than it was last year. The three fifth-grade teachers have divided the curriculum among themselves, with one teaching social studies, another teaching mathematics, and the third teaching language arts, arts and humanities, and practical living/vocational studies. Each fifth-grade teacher is responsible for teaching science to his or her own homeroom.

The school uses Reading Renaissance[5] in its 3-4 and 5 classes and received district incentives for doing so. The STAR[6] testing program accompanies the reading program.

Willow Middle School

Willow County Middle School is a large county school located in a mid-sized town, which also has an independent school system. Willow County Middle School has a relatively low percentage of students who qualify for free or reduced price lunch and a very low percentage of minority students in its student population of more than 700. The principal described most of his students as coming from stable backgrounds, the exception being a group of students who move frequently and who make up a substantial portion of the special education population. About five years ago, a new middle school was added in this district, and the principal said that the new school took a portion of his school's wealthier students.

The principal said that his staff and he believe the CATS system is more suited to their school than was the KIRIS system. He described the Kentucky Core Content tests as taking a more

[5] Reading Renaissance (www.hanlundphillips.com/fetc/EE.htm) is a reading program designed by the Institute for Academic Excellence. It uses the STAR testing system designed by Advantage Learning Systems, Inc. as a monitoring tool. It is designed to assist teachers in promoting, instructing, monitoring and intervening in the reading classroom.

[6] STAR reading tests (www.advlearn.com/starreading/default.htm) are computerized tests that determine student reading level. They can be used throughout the school year as a diagnostic tool for assessing the progress of individual students in reading. Advantage Learning Systems, Inc. produces them.

traditional approach, and said that they were more computational, less conceptual, and more fair. This was a vast improvement over the previous tests, he said. The school's scores improved, as well, showing strong growth in writing. He said that they were not happy with their science scores, however, and had emphasized science as a result. Their school board had given more money for science equipment, for example, and they were requiring more hands on or laboratory work in science classes.

The school also had switched from a 7-period day to a flexible block schedule. Teams could decide how to use the time, but they were required to turn in a weekly report to the principal detailing how they covered the Core Content topics.

The school was participating in the Accelerated Reader program, and the first 30 minutes of each day were spent in that program. They also emphasized writing across the curriculum and did open-response questions each week as well as practicing how to interpret multiple-choice questions.

Cypress County District

Cypress County is a South Central Kentucky school district and serves a very rural community. All the schools are located in the county seat, which is a small/average size town. The outlying county schools were consolidated in the past 5-6 years to a central campus area. The central campus area has a new high school building and the remaining schools occupy older buildings. Many of the older schools contained a newer section built more recently to contain the increased population of students. Instead of dividing the schools into several separate elementary or middle schools because of the large population, each of the schools houses fewer grade levels than is typical. The district has a primary school (K-1), an elementary school (2-4), an intermediate school (5-6), a junior high school (7-8), and a more traditionally configured high school (9-12).

This method of splitting students into classes and schools has both advantages and disadvantages. First, the Kentucky Core Content Test for the elementary grades is given at the fourth and fifth grades. This means that the test is split between two schools in Cypress County. One school is tested in science and reading while the other is tested in social studies and mathematics. Despite the proximity of the schools, the communication between teachers from different schools was not sufficient for the two schools to operate with a unified effort toward assessment.

Cypress Elementary

Cypress Elementary is the grade 2-4 school. It is currently being renovated and much of the interior of the school is a construction site. The school is not set up to be very team-friendly, with isolated classrooms rather than the more common pod design, and not much evidence of real teaming was observed. In fact, the most collaboration seen at the school was between teachers who were not on the same team. The school is relatively new to the team concept and most teachers seem to be continuing with the more typical self-contained pedagogy. There are at least five teachers at the school who move from classroom to classroom and collaborate with the

homeroom teachers. They include a writing specialist, a science specialist, an art teacher, and at least two physical education teachers.

The school walls were covered with student work. The projects evidenced on the walls of the school indicate that the students spent a great deal of time completing their assignments. Many of the projects obviously took at least several days to complete. In addition, the work that the students performed in observed classes often took longer than a single class to complete and seemed more challenging than the work in many of the other schools we visited. The lessons themselves were often very traditional. We saw spelling lessons, note taking, and lecture during our visit.

This school can be characterized further by its response to a good showing on the accountability test. The principal and teachers have accepted responsibility for student test scores and are proud of their accomplishments. The are cautiously optimistic about future tests. They have identified some strategies that they believe are successful in preparing students for the assessment, and they are striving to scale up the implementation of those strategies. Their major concerns regard the primary grades, where the lack of accountability is seen as a weakness of the system. The influence of the primary grades on their school's performance is viewed as a limitation of their responsibility and their potential success.

Cypress Intermediate

Cypress Intermediate contains the 5th and 6th grades in the Cypress County District. As such, Cypress's only direct contact with the Kentucky Core Content Test is in 5th grade mathematics and social studies (also arts and humanities and practical living/vocational studies). Scores in mathematics were lower than expected last year and scores in social studies were perceived as unrealistic due to the high standards evidenced by the statewide percentile scores. The result for Cypress is a sense of urgency to bring up math scores. The school has implemented a supplemental mathematics basics program. Every teacher at the school, including the principal, spends 45 minutes working with a few (7-8) students each morning. They hope that this remediation will fill in the gaps in the students' previous mathematics instruction.

The need for remediation, as opposed to supplemental mathematics instruction, stems from an assumption regarding the elementary school. Because the elementary school is tested in reading and science, it is assumed that those subjects take precedence in grades 2-4. There was certainly some evidence to support emphasis of those subjects at the elementary school, but clearly math and social studies were not omitted from the elementary curriculum. However, because the two schools serve exactly the same population, differing student test scores cannot be the result of differing demographics. Because communication between schools is poor, the search for reasons for the undue differences in scores has led to rumor spreading and animosity. If one school does better than the other, rumors abound of cheating the system or sabotaging the long-term curriculum for quick gains on the assessment.

The need to improve test scores has caused Cypress to become much more overt in its attempts to promote changes in instruction. Open-response questions were required on every test this year and were to be turned in to the main office for verification. Student work was posted,

although lacking in quality (compared to the elementary school), on most of the classroom walls. Professional development has emphasized open-response techniques and the creation and use of scoring rubrics.

Cypress Middle School

The middle school contains the 7th and 8th grades. The facility has been added on to over the years and the current floor plan spans two separate buildings. The older building, which was once the high school, contains classrooms, the main office, and the gymnasium. The newer building contains the cafeteria and another set of classrooms. The students were not divided such that all their classes were in one building, so security was a concern for the school during class changes. The classrooms themselves were covered in student work; however, most of the student work was of low quality compared to other schools in the study and did not evidence the same effort as the student work displayed at the elementary school.

Perhaps the most memorable portion of the visit was hearing about a sex education program the school had recently adopted. In cooperation with the high school, the school has begun a program aimed at convincing students to remain abstinent during their adolescence. The students learn about human reproduction in science class and then attend a program that combines discussions of sexually transmitted diseases held by local physicians, fun activities (dating etiquette, attraction, etc.), and small group discussions with high school students. The program's existence is somewhat controversial given the conservative nature of the community. One of the teachers explained that he was nervous about meeting with a parent because of the nature of what the students in his class were studying.

Another program that distinguished this district involved frequent visits from district personnel. The program was district-wide, but the focus seemed to be on this school. Briefly, district persons visited the school, observed classes, and questioned students. Their questions were fairly straightforward and included, "1—What are you learning? 2—Why are you learning it (what use is it in the real world)? 3—How do you know if you are doing well?" Students were chosen to answer these questions individually and were expected to know the answers. Teachers were making the answers to these questions very overt in their lessons. The principal explained that the visitors could show up at the school at any time and provided no warning for the school.

Discipline problems at this school were similar to those found in most large rural middle schools in this study. The principal focused on stories of problem teens and the variety of the population the school served when we spoke with him. He seemed to be the person primarily responsible for enforcing discipline at the school.

Dogwood County School District

Dogwood County is a reasonably rural district in Southeastern Kentucky. The county contains several elementary schools, but all of them feed into a single large middle school. The middle school is located near the county seat, which is a medium sized town. Many of the elementary schools, however, are very isolated. Several of the teachers we interviewed had attended the same school in which they now taught and had gotten their teaching credentials through

the small college located in the county.

The county was fairly new to the idea of curriculum alignment across grade levels. They began implementing an alignment plan called the Core Knowledge[7] two years ago, which seems to be a curriculum framework, similar to the Core Content. Evidently a good deal of effort was made determining the degree to which the Core Knowledge matched the Core Content and incorporating the two together.

Dogwood Elementary School

Dogwood Elementary is an old school located in a particularly poor area in the county. The front door of the school opens into the gymnasium. Classrooms surround the gym, which is noisy when it is in use, and the noise carries into the surrounding classrooms. Opposite the front door, stairs lead down into a newer section of the school. More classrooms, the library, the cafeteria, and the main office are in the newer section. Conduit runs along most of the walls of the school, evidencing new phone lines and updated electrical service.

Teachers at the school were pleasant, although they seemed a little nervous about having visitors in the building. The number of students per class was very different for different grades, with some grades having as many as 30 and others as few as 18. The teachers accepted this as the status quo, although no one gave a compelling reason for the disparity.

Computers and other evidence of technology were scarce at the school. Evidence of prosperity (such as new desks, teaching materials, etc.) was sparse. Evidence of prosperity among the students was also largely absent.

Instructional practices at the school were very similar to those seen in most other Kentucky elementary schools in this study. Teachers typically used very reform-friendly (cooperative learning, inquiry, group participation and discussion, etc.) practices. Students were mostly on-task and diligent in their work. The difficulty level of the lessons was comparable with other schools. Perhaps the best way to describe the school is to say that it is an average rural Kentucky school with a much higher than average proportion of impoverished students.

Dogwood County Middle School

The county middle school contained students from all the elementary schools. One teacher explained that the county was divided into wealthy and poor sections, so while there was not the concentrated poverty of Dogwood Elementary, the disparity of wealth among the students in attendance came with its own set of challenges. One of those challenges was the perception among the middle school staff that students receive very different elementary instruction depending on which school they attended. Calls for common textbooks and a unified curriculum were common among middle school teachers. One explained, "They may not have had any science instruction in

[7] Core Knowledge is a curriculum alignment program created by the Core Knowledge Foundation. It stresses consistent aligned teaching toward a shared curriculum. More information can be obtained at http://www.coreknowledge.org.

the past two years when they get to me. Then I'm responsible for bringing them up to level before the test."

The school building itself was set up in pods. The seventh and eighth grades were very separate in the school, each with its own wing of the building. Each pod was occupied by one team, which included one social studies teacher, one language arts teacher, and one math teacher. The teams also had a science teacher, but science classes were taught in laboratory classrooms that were in a separate pod. Another pod housed a shared set of computer labs.

This school was very overt in its attempts to perform better on the Kentucky Core Content Test. Their professional development opportunities were centered on content coverage/curriculum alignment and improving students' writing abilities. The Core Knowledge materials hung on the walls of most classrooms and topics were checked off as they were covered. The school also seemed to be experimenting with various academic programs, but hasn't had enough positive results to gauge what is and is not effective. The principal described several programs that had been started over the past 2 years. The process of deciding what constitutes improvement at Dogwood County Middle was not dissimilar to the process many other schools participating in the study have undergone. They just seemed to be a couple of years behind many of the more successful schools.

Participant Demographics

Table 4 lists the total number of elementary and middle school teachers interviewed during the study. Table 5 contains data regarding the average number of years teaching experience reported by the participating teachers.

Table 4. Teachers Interviewed During Study

	Total	*Science*	*Social Studies*	*Other*
Middle Schools	127	62	61	4
*Elementary Schools**	101			

*Note: Elementary schools rarely had teachers who taught only science or social studies. Attempts to categorize them as such would be misleading and are omitted from this table. Middle school numbers refer to teachers who taught at least one science or social studies course.

Table 5. Average Number of Years Teaching Experience

	All Teachers	*Science Teachers*	*Social Studies Teachers*
Middle Schools	10.9	10.2	11.5
*Elementary Schools**	12.1		

*Note: Elementary schools rarely had teachers who taught only science or social studies. Attempts to categorize them as such would be misleading and are omitted from this table. Middle school numbers refer to teachers who taught at least one science or social studies course.

Findings

This portion of the study is divided into six sections. Sections represent major themes or topics that emerged from the data collected. Each section contains subtopics that more specifically represent the information provided by participants.

The Influence of CATS

Not Much has Changed Since Last Year

The first important finding from conducting this research was that schooling was continuing in much the same manner as last year in the 31 participating schools. Widespread programmatic changes have not occurred as a response to the changeover from KIRIS to CATS. Schools were continuing the practices with which they had enjoyed success and were making minor adjustments in much the same way they made adjustments when they received KIRIS score reports.

Schools were still more likely to try to repair perceived weaknesses than to build upon strengths. Schools examined the data they received via the Kentucky Core Content Test reports to see what subject was lagging behind the others or what sub-content topic was declining. Each weakness that was identified was attacked via professional development, program adoption, remediation, hiring new school staff members, schedule alterations or other means. However, once the weakness was supplanted in the data schools received, the methods that were used to correct it were aimed at another target.

Students continue to spend a great deal of their school day writing. Often they keep class logs or journals. They are sometimes required to write reflective works related to the previous day's lesson. When they work in groups, the groups typically hand in a lab report or some other written product. The reduction in portfolio requirements has not greatly diminished the emphasis placed on writing at the participating schools. Teachers report that they continue to spend a great deal of time in the preparation of portfolio entries. Most schools have seen minor increases in the use of multiple-choice tests in the classroom and class time devoted to teaching multiple-choice strategies.

The core subjects, language arts, mathematics, science, and social studies, still make up most of the curriculum at most schools. The Kentucky Core Content for Assessment remains the document most cited in curriculum planning. Arts and Humanities and Practical Living/Vocational Studies were receiving increased emphasis at many schools. Curriculum alignment efforts within schools and districts continue, although some districts were only beginning the process while others have comprehensive alignment plans in place.

Teachers' Impressions of CATS

Most of the teachers we spoke with had positive things to say about the changes in the testing system. They mentioned specifically the addition of the multiple-choice questions to the calculation of the schools' accountability indexes and the additional subcategories within the Novice and Apprentice performance designations. It is important to note, however, that the improvements in the testing system were typically considered minor and their recognition by

teachers does not represent a high level of support. Teachers typically couched their statements by adding "It's a step in the right direction," or "It's a little better than before, but…"

Teachers remain skeptical concerning the testing system. While the shift from KIRIS to the Kentucky Core Content Test dealt with a good number of teachers' concerns about testing and accountability, others have not been sufficiently addressed. Many teachers still harbor concerns about testing different cohorts of students each year, the effects of population demographics, test administration procedures, and other issues that limit their support for the new system.

Other teachers were quick to say that they were misled into believing the Kentucky Core Content Test would be new. "We were promised a new test, but its just the same old test with a different cover," was a common protest. One teacher from Dogwood quipped, "They think we're stupid and that we wouldn't notice that it's just the same old test." The notion that KDE had deceived schools into believing in changes in the system that were not genuine was another common theme when teachers discussed the new test.

A smaller group of teachers claimed that the test was improved considerably for reasons that are not supported by the changes in the system. At least a few of the teachers claimed that the test was better because the questions were more specific and more focused. They evidently did not realize that all of the "live" items on the test were included on the previous KIRIS assessment. The item pool for the test was the same except for pretest items that do not count toward student scores or school accountability indexes at all. It seems that for a select minority, the belief that the test was new was sufficient to cause the attribution that it was also improved.

Teacher Practices

Teacher practices have not changed a great deal since last year. In fact, many teachers with whom we had spoken previously did not bring assessment materials to the interview because they had given them to us last year and had not changed them. Changes the teachers did describe were often continuations of programs that were begun last year or earlier. An exception to this rule was the multiple-choice strategy lessons that many teachers had added to their curriculum this year.

Classroom Practices

Reading and Writing

Most of the schools participating in the study had some program in place to promote better student reading and writing. Some had hired special teachers who focused on remedial reading. Others had dedicated portfolio workshops where students were either peer-tutored or simply given additional time and resources to work on portfolio entries. At least two of the participating schools had more language arts teachers per student than teachers per student in other classes. This allowed for smaller language arts classes where students could receive more individual attention. A few of the participating schools also had more class time devoted to language arts than other subjects to allow the teachers to conduct individual student conferences and to give

students more time for completing portfolios. Many of the schools had adopted the Accelerated Reader program.

The reasons behind the creation or adoption of these programs were really very simple. Schools count on the test being skewed toward reading and writing skills. Numerous principals and teachers explained that it does not matter how good a math student a child is if she can not read and understand the questions and express her thoughts on paper. Some teachers expanded this line of reasoning by claiming that the test was biased toward good writers. The requisite reading and writing skills are seen as a barrier for many students who "know the right answer, just not how to answer it." Low scores in other subjects were seen to reflect this lack of reading and writing skills rather than a deficiency in mathematics, science, or social studies.

Portfolio Requirements

Nearly all of the teachers we interviewed during the course of this study cited the reduction in portfolio requirements as a positive change in the testing and accountability system. Some teachers were adamant that the portfolios should be eliminated altogether. The most common complaints about the portfolios included the time factor, which many teachers viewed as time taken away from content instruction, and the reliability of grading the pieces once they were complete. Teachers pointed out the biases each reader must necessarily bring to the evaluation process.

Perhaps surprisingly, several teachers at the elementary level told us that they still used portfolios in mathematics. They spent less time on them now compared to when they were evaluated, but they still required students to complete them. Many of these teachers valued the portfolio as a teaching tool but not as an accountability component. They claimed that they spent much less time polishing the portfolio pieces in class, that the pieces were generally much shorter than before, and that the review processes they used were not nearly as extensive as before. Many other teachers, however, stopped using the mathematics portfolio as soon as it was dropped from the testing and accountability system.

Despite the reduction in the number of writing portfolio pieces required for each student, most teachers explained that the amount of time spent completing the pieces had not been appreciably reduced. There were essentially two schools of thought that have sustained the time factor for completing portfolios. The first was that the more time students had to work on portfolios the better they would be. Teachers who believed this spent the additional time they would have used completing another entry in polishing and editing the pieces that were required. Students spent the same amount of time; they just took longer to complete each entry. The other likely reason the time factor has not been reduced was that several teachers use a "best evidence" system, where each student chooses the best work from several potential portfolio entries done throughout the year. These students simply chose one less entry, leaving the time allotted for this work essentially unchanged.

Portfolios Versus On-Demand Writing

Many elementary teachers expressed that they had not had much success in raising on-

demand writing scores. They often explained that they were spending more time writing in class than ever before, but the change in scores did not reflect the changes in emphasis. Several school-wide initiatives designed to improve reading and writing were described very positively by these teachers, but the common perception was that those initiatives were improving reading scores, portfolio scores, and perhaps other content area scores. Preparing students to perform well on the on-demand section remained mysterious for many of the participating teachers.

Many teachers interpreted low scores on the on-demand section of the Kentucky Core Content Test and high scores on portfolios as evidence of a lack of accuracy in grading both sections. The same criteria were used to judge both types of writing, although the preparation time for the on-demand writing was necessarily much shorter. They explained that if their students could write well on one form, then those skills should carry over. They elaborated that they worried about scoring accuracy, scorer bias, regional interpretations of the questions that could lead to lower scores, and other concerns. In some schools a high on-demand score compared to portfolio scores was interpreted as a lack of diligence in preparing the portfolios. In others, however, there was an apparent expectation for standards to be norm-referenced and to be uniform across all tests, including portfolios. Roughly the same proportion of students were expected to score proficient on the on-demand section of the test as on the portfolios. Because the on-demand scores were typically lower than portfolio scores, either the test was considered inaccurate or the apparent lack of progress on the on-demand section was interpreted as an indication of the state's unreasonable expectations. Several of the teachers expressed that they had very little confidence in the on-demand writing scores.

Multiple Choice

Instructional emphasis was still clearly focused on writing, but many schools reported using more multiple-choice questions on assessments and for regular class assignments. Most teachers were quick to point out that they had never completely abandoned multiple choice as a method, but that they had reduced it more and more in the years after KERA was implemented. One principal joked during Phase 1 of this study that the Scantron grading machine the school had purchased in 1990 was essentially unused until recently (Thacker et al., 1999).

Teachers who eliminated or nearly eliminated multiple choice in their classrooms explained that their students didn't have any experience answering those types of questions. Many of them taught a special "how to answer a multiple-choice question class" last year and have plans to do so this year. The classes typically included practice limiting choices and some testing strategies for questions that were completely unknown. Always guessing the same letter, reading all answers before choosing one, and other simple strategies were being reintroduced in these classrooms.

Breadth of Core Content Requirements for Science and Social Studies

A consistent concern among science and social studies teachers participating in this study, as well as those participating in previous studies (Thacker, Hoffman, & Koger, 1998), was that the Core Content for Assessment is too broad to effectively cover. Teachers reported that they were forced to choose between breadth and depth of coverage when they planned their curriculums.

This balance is a continual dilemma in the design of any curriculum, however the high stakes accountability system coupled with a reasonably prescriptive set of curriculum guides make it especially troublesome in Kentucky.

Science teachers pointed out the large number of topics included in the Core Content for Assessment as an indication of the overwhelming scope of the curriculum. They claimed they felt responsible for teaching biology, chemistry, meteorology, physics, astronomy, ecology, and other sciences to students in a single year while any one of those topics might be better suited to its own course or group of courses. They responded to these perceived responsibilities by either teaching a few of the topics and hoping that the test content matched their curriculum guesses, or by teaching only a surface level overview of all the topics. Some teachers pointed out that curriculum alignment efforts had ameliorated this dilemma, but that it was still a major concern.

Social studies teachers were confounded by the scope of their curriculum in another way. Most had received training in only one of the five areas comprising the social studies core content. They find themselves responsible for teaching history, government, economics, culture, and geography when they might have been trained to teach only one or two of those topics. Many social studies teachers in this study reported that they simply didn't have the time in class to teach students the history of the particular timeline they had been assigned and incorporate all the other aspects required by the core content.

Arts and Humanities and Practical Living/Vocational Studies

Although not a focus of our interview questions, issues related to Arts and Humanities and Practical Living/Vocational Studies regularly emerged during our interviews. Several participating schools described a large emphasis on Arts and Humanities and/or Practical Living/Vocational Studies. In the current system the two tests each have half as many items as the other subject tests and all the questions count. Together they account for roughly the same portion of the school's accountability index as one "core" subject.

The Arts and Humanities test contains questions referring to dance, music, artwork, theater, architecture, philosophy, and other topics typically included in discussion of the Arts or the Humanities. The Practical Living/Vocational Studies test contains questions referring to health issues, career planning, budgeting, sports, personal responsibility, safety and accident prevention, and other topics. These tests are given to students in the 5^{th} 8^{th}, and 10^{th} grades.

Most schools did not teach classes specifically devoted to Arts and Humanities or Practical Living/Vocational Skills. Instead, students were expected to learn about those topics in their regular classes, through extra-curricular activities, outside school, or through classes centered on one of the topics included under the umbrella of Arts and Humanities or Practical Living/Vocational Skills. Those classes were typically not taught as frequently as the core subjects, but made up a substantial portion the total curriculum for many students. They included band/music, art, health and physical education, and others. Many of those classes were not required and were only taken by a small portion of the student body. Specialty teachers, who may not be full-time faculty members, often taught them. One school was very proud of the art teacher that they shared with another school in the district. The art teacher spent about three days per

week at the school and rotated through the "regular" teachers' classrooms as well as leading an after-school art program. Student paintings were prominently displayed in the lunchroom of the school.

Several participants in this study related that they felt that they had done all they could to improve test scores in the past. Schools looked toward Arts and Humanities and Practical Living/Vocational Studies as a "new area." While student scores in the other subjects may have plateaued despite the best efforts of the staff, there was an expectation that scores on these subjects would improve if similar measures were taken. Toward that end, we have seen schools that have decorated their walls with famous works of art, schools that have posted dance movements in their classrooms, schools operating small businesses with students as employees, elaborate art studios, music rooms, student-operated television stations, or vocational/technical workstations (separate from the more common computer rooms). The amount of resources and the expenditure of energy required maintaining the programs that correspond to these school assets were often extreme. The return on these investments in terms of test scores was largely unknown, but it seemed clear that many schools expect their scores to improve due to these efforts.

Professional Development

Professional development, or in-service training, is one of the few avenues that schools have for altering instructional practices among teachers who are already in the classroom. When teachers were asked whether their schools had created any professional development opportunities in response to the Kentucky Core Content Test scores they typically responded in the affirmative. This was not surprising given previous research on the relationship between KIRIS and professional development (Borko, Elliot, & Uchiyama, 1999; McDiarmid, David, Kannapel, Corcoran, & Coe, 1997; Thacker, Koger, & Koger, 1998). There was, however, a subtle shift in the type of professional development activities the teachers at many schools described.

When teachers described their professional development experiences in a previous HumRRO study (Thacker, Koger & Koger, 1998), they would almost invariably describe the four-column method, power verbs, writing and/or assessing open-response questions, or other similar programs. These programs were designed to improve student test scores by making the students better able to respond to the format of the test. They cross all subject areas and were perceived to be beneficial for all teachers. Many schools in this study were still focusing their attention on this type of program. Others have shifted their emphasis to be more content specific and seem more interested in the content of the test than in its format.

Fairness Issues

Accountable Grade Teachers

The Kentucky accountability system was designed to work at the school level. Education was assumed to be cumulative and all teachers were assumed to contribute to the eventual composite accountability index irrespective of the grade they teach. The pressure to improve was supposed to be spread throughout the education system through a rewards and sanctions/assistance

system. Presumably, each teacher at a school was expected to contribute to the overall betterment of the instructional quality and the eventual test scores of the students at the school.

These expectations are not played out in exactly this manner in most public schools however. Teachers in the accountable (tested) grades have been under a tremendous amount of pressure to bring up test scores (Wolf & McIver, 1999; Wolf & Borko, 1999), while teachers in other grades could afford to take a more distant view of the accountability system.

A common concern among teachers in accountable grades was that students come to them unprepared and unable to perform at grade-level. Many reported that their students arrive with poor skills and poor vocabulary. A rigorous instructional program, designed to improve test scores, was beyond the experience and ability of many of their students. These teachers were, therefore, forced to take time from their planned curriculums for remediation.

Some schools were much more overt in the ways that they pressured the accountable-grade teachers than others. Cypress County District performed a secondary analysis of the data by individual teacher. They used the data to create performance ratings for each teacher that included an overall rating and level of improvement. They used the subscores to determine how the teacher should focus her professional development time. School officials also used the data to focus assistance and scrutiny of individual teacher's instructional practices.
The Cypress district office also put pressure on accountable-grade teachers to improve. They performed surprise inspections of classrooms and asked students questions regarding the value of their assignments. If the students gave unsatisfactory answers to those questions, even more pressure was brought to bear on the teachers.

Cohort Effects

The phenomenon of the atypical class was still seen as a threat to the validity of CATS by many of the participating teachers. Teachers frequently told of students in the past years that just "got it" quicker than the current students. The practice of setting baselines and determining goals based on those baselines was seen as a weakness if you happened to have a "gifted" baseline group. The teachers who knew that the baselines were being reset based on the interim accountability cycle were skeptical, especially if their particular group of students were performing at or above expectations.

A few principals and teachers explained to us that it might be advantageous to set a low baseline score deliberately during this cycle. The penalties for low performance have been greatly reduced and having a low baseline would seemingly make gains easier for the subsequent years. All of these participants explained that they were not planning to sabotage the testing system in this way, but often claimed that they were worried that other schools might.

Many of the study participants viewed whether a class was low performing or high performing as a phenomenon of the class rather than the teacher. Luck determined if the school set a high or low baseline. Luck, by association, was also viewed as controlling the likelihood that a school would improve or decline. The cohort effect, whether real or imagined, was a serious issue for many teachers and it remains a threat to the perceived validity of the testing system.

Standards Seem Inconsistent by Subject

Standards in science and social studies were a major concern with many of the teachers and principals who participated in this study. Their concerns stemmed primarily from the fact that very few students score in the top two proficiency categories in those subjects compared to mathematics and reading. Less than 1 percent of fourth- and seventh-grade students scored Distinguished on the Kentucky Core Content Test in science. Less than 6 percent of fourth-grade science students were classified as Proficient and Distinguished combined. In seventh grade the total percentage of students scoring in the Proficient or Distinguished categories combined was less than one half of 1 percent. Less than 1 percent of fifth- and eighth-grade students were classified as Distinguished on the social studies test. Nine percent of fifth-graders and 12 percent of eighth-graders were classified as Proficient on the social studies test. School staff members were quick to point out that the sub-categories within the Novice and Apprentice categories were a positive change in the system, but they remain convinced that the standards for science and social studies are unreasonably high despite the additional points now available.

Several teachers argued that they didn't believe scores in science and social studies would ever reach an average Proficient score. In both elementary and middle schools, scores in those subjects have not improved at as great a rate as in mathematics and reading. The contributions of teachers in those subjects toward school-level rewards and sanctions/assistance were often ambiguous or poorly articulated. This seems to have led to a backlash in at least one school where attention paid to those subjects has waned in favor of subjects that were perceived to be more likely to raise school scores. The eighth-grade science teachers described themselves as the "red-headed stepchildren" of the accountability system. They were instructed to spend a large amount of their class time teaching topics associated with arts and humanities or practical living/vocational studies. The classes were science in name, but their curriculum had more to do with health and other tangentially related topics.

In another school, a seventh-grade social studies teacher explained that his curriculum had suffered recently because of the emphasis on language arts at the school. He mentioned that the language arts teachers had fewer students per class and more time was allotted for each language arts class. A larger portion of school funds was devoted to purchasing language arts supplies and for funding professional development in language arts. Social studies, mathematics, and science teachers were being asked to add the arts and humanities and practical living/vocational studies curriculum to their classes despite complaints that the science and social studies curriculum is already too broad to teach effectively. The principal at the school explained that the reasoning behind the emphasis on language arts was a response to the format of the test. Since the test has a large open-response component, language arts was perceived to transcend subject area. Strong language arts skills were seen as a means to improve all subject scores and the ability to write seemed more important at the school than content area knowledge.

Standards and Fairness Across Grade Levels

Many teachers were also quick to point out that the standards were apparently not the same across grade levels. They meant by that statement that the percentages of students in each

performance category were very different in elementary, middle, and high school. Middle school teachers were most vocal about the disparity. On teacher explained, "They didn't suddenly get dumber when they got to middle school and then get smarter again when they left. The differences are in the test, not in the kids."

Technology/Resource Equity

The Kentucky Education Reform Act was implemented due to disparities in funding among Kentucky's public schools. KERA was designed, in large part, as a method of "leveling the playing field." Since the beginning of the reform effort, the funding gaps between the state's wealthiest and poorest schools have been greatly reduced. The amount of funding available for school improvement is at an all-time high in Kentucky (Cody & Guskey, 1997). Professional development funding has improved more than twenty-fold (McDiarmid et al., 1997). The technology initiatives have allowed nearly every school in the state to access the Internet, word-process portfolio entries, and many have even established their own web-sites. It seems clear that Kentucky's schools are financially in better shape than they were before KERA.

Many of the disparities, however, remain. During the course of this research we visited a school with an Internet connection for roughly every 2 students in the building. We also visited a school with a severely leaking roof and very little of the obvious technology emphasis seen at most Kentucky schools. How disparities such as this one affect students' opportunity to learn is unknown.

This is not to say that the gap between richer and poorer schools in Kentucky hasn't been reduced. It merely illustrates that the differences between schools have not been completely eliminated. Since the disparity between students' opportunities to learn led the original system of education in Kentucky to be declared unconstitutional, it would seem prudent to continue monitoring the realization of students' opportunities at representative schools.

School Sanctions

Schools awaiting sanctions for testing violations were unsure about their fates. At least two of the schools participating in this study had violated testing procedures. They had reported their error and were awaiting a decision regarding their scores. The principals at these schools were unsure when they would receive a decision, what the possible ramifications of that decision might be, what penalties were likely, or even how that test administration penalties would affect baseline scores. One went so far as to comment that the penalties might actually be good for the school, since a lower baseline would make it much easier to appear to have improved.

Test Score Analysis Issues

Sub-content scores

Each Kentucky public school received a set of data regarding the calculation of their accountability index. It included student test scores, proficiency categories, and percentile rankings from the Kentucky Core Content Test. In addition to those measures, sub-content scores were also included. Sub-content scores refer to student performance within each subject-level test. If a domain is defined as social studies, there are several sub-domains under that heading. For instance, the social studies test might include sub-content scores for culture, economics, history, government, or others. This year sub-content scores also included scores for sections within each sub-domain. These scores were even more specific. They fall under the other sub-categories. For instance, government might have a section on county, state, and federal government.

These scores were produced at the school level. Students only take one of the six forms of the Kentucky Core Content Test, and results from all forms were required for this level of detail. Arts and Humanities and Practical Living/Vocational Studies were not reported at the section level due to the smaller number of questions on those tests, but were reported by sub-domain. The reliability of sub-content scores is necessarily less than the reliability of the proficiency categories due to the smaller number of questions they rely on and the smaller number of students at any one school who might have answered the questions on which they are based. The validity of those scores depends, of course, on the use of the scores at the school and classroom levels.

Score Analysis and Instructional Practice

Teachers very rarely described changes in their teaching pedagogy that had come about as a result of the switch from KIRIS to CATS. If their school had enjoyed some success in the system, they reported that they were continuing with those practices that had proven to be effective. If their scores were not as high as they felt they should be, the reasons were typically described as beyond their control. Student population, parent apathy, and poor preparation in the lower grades were all examples of the external factors that teachers reported as limiting their ability to make the kind of improvements called for by the CATS system. While this study was not designed to estimate the extent to which those factors influence student scores, it demonstrates that the displacement of control surrounding student scores has limited the changes in pedagogy attempted by the teachers at some schools.

The changes in practices that teachers did describe were subtle and typically appeared to be short-term responses to testing format changes rather than changes representing a shift in pedagogy. Many teachers have begun teaching students strategies for answering multiple choice questions. Those lessons were typically scheduled near the testing window and usually only occupied a couple of class periods. They were often described as one means of examining student results on scrimmage tests.

Changes in teaching content as a response to CATS were usually described in terms of the sub-content scores. A typical response from science teachers asked about changes in their

teaching practices related to the test was, "I'm emphasizing life science more this year because our scores weren't as good in that area. We've always done well in life science until this year, and the kids seem to grasp it easier than the other subjects, so I don't understand how we could have dropped in that area."

Score Analysis and Professional Development

Several teachers explained that the sub-content scores were used to plan professional development as well. They described the contribution of the sub-content scores to the construction of the Consolidated Plan and their own individual professional development plans. For instance, because one school's economics sub-content score had dropped, teachers from all areas were sent for training in teaching economics. The school then required that each teacher spend part of the year teaching economics in their classes.

Sub-content scores and Program Decisions

Some of the participating schools had discontinued programs based on student performance in a sub-content area. Despite teacher reports regarding the high effectiveness of some programs, their continued existence was threatened by a decline in the sub-content score most closely associated with the program. For instance, a school explained that an elaborate integrated program on the various world cultures, culminating in a "cultural fair," supported by students, teachers, parents, and community members was unlikely to be continued because scores on "culture" went down this year. The attribution made by the teacher was that the students had not gotten a broad enough exposure to all the cultures, but had instead concentrated primarily on the one that they researched, displayed, and presented. The program will likely to be replaced with a broad overview of the several cultures studied during class time.

Earlier Score Release

Receiving the test scores earlier in general, and the sub-scores in particular, was perceived as an improvement in the overall testing and accountability system. Teachers reported that they could alter their curriculum, adjust their patterns of emphasis, and adjust the order in which topics were covered in the classroom prior to the next administration of the test. Many expressed that they would try to address deficits in student knowledge prior to the test administration.

Test Score Interpretation

The changes in the accountability and testing system have caused the state to issue serious cautions regarding the comparability between 1998 KIRIS scores and 1999 Kentucky Core Content Test scores. Many schools have therefore limited their data analysis to comparison with the state average. This level of analysis is certainly appropriate and is not a new practice. Many schools reported that they had always used the relationship of their own scores compared to the state average in their analyses. Some schools also compared themselves to other schools in their regions, but were often not sure how to interpret the results.

Conversion of Scores to the KIRIS Scale

The need for data from which to plan or, more importantly, the need for data to show that your strategies are "data driven" is required in Kentucky. Schools are required to produce a Consolidated Plan for improvement as a part of KERA. In the Consolidated Plan are strategic initiatives designed to improve the educational opportunities and therefore the test scores of students. It is not enough that those initiatives are soundly reasoned, they must be data driven, and the data that drive the changes must be identified in the text of the plan. Schools have come to rely on test data to serve these purposes.

The test shows strengths and weaknesses by subject, and through the sub-scores, at even more specific levels. By examining trends in student scores, schools can choose to continue, alter, or eliminate programs. In addition, schools use the scores to justify the expense of programs that are purchased from external sources. Whether the use of this data for these purposes is justified or not, the system has created a reliance on the data for meeting its own requirements. As one teacher jokingly put it, "They tell us not to compare the data from this year to last year, but then they require trend data as part of the Consolidated Plan. Which should we do, make the comparison or not?"

Perhaps it was this need for data, which led some districts to create a mathematical formula to calculate what each school's index would have been under the old system. Schools then compared scores from last year to this year and determined if they had improved, declined, or maintained their standing on the test. This data was then used to determine trends by subject.

The actual calculation used to determine what the scores might have been on the old test were not available, so no judgments regarding their accuracy were possible. However, the accuracy of these calculations is suspect due to the difficult nature of scaling and equating scores from year to year and the various changes in the testing system. The changes were so severe that the state did not perform (or require of the testing company) the same kind of linking and equating as in the past. Instead they chose to use an equipercentile formula to create cut points for assigning proficiency categories to Arts and Humanities and Practical Living/Vocational Studies. The rest were equated using the multiple-choice items. Scoring drift was not analyzed for open-response items. The Kentucky Department of Education also issued warnings that the two tests were different and were not comparable (KDE, 1999a).

Early Release of Scores—Impact for Individual Teachers

Nearly all the teachers interviewed reported that receiving the test scores earlier was a positive change in the testing system. "The earlier the better," was a common theme among teachers and principals. When asked for reasons that receiving the test earlier was helpful, however, answers were often vague and rarely depicted any concrete changes in the curriculum or teaching practices. "We can plan what to do earlier now," or "We know what we need to work on earlier," or "It was good to get the scores earlier as validation that we're doing a good job," were typical answers to this question. Follow-up questions rarely resulted in more specific answers.

Although this study occurred at least one month (at most three months) after the test scores

were released, one school still had not released the results to its teachers. What little information they had came from the newspaper. The principal at the school explained that the scores were still being analyzed at the district level and would be explained to teachers during an upcoming staff meeting. Other schools explained that they were still performing analysis, but they had released enough information to allow teachers to gauge their progress.

Diagnostic Use of Test Scores

One of the major advantages of schools receiving test scores earlier was that the scores might be used to diagnose deficiencies among individual students. This was especially true of fifth- and eighth-grade students, since those students were typically still attending the school that received their fourth- and seventh-grade results. So, every fifth- and eighth-grade teacher received at least science and reading scores for their current students.

Only a few of the teachers participating in this study reported that they had used the data in any meaningful way. Many had not seen the scores of their current students at all, and a number of teachers reported that they did not want to see their students' results. When asked why they didn't want to see the scores, they explained that knowing the scores might limit their expectations for individual student performance. Pre-judging a student might cause them to settle for work that was not the best the student could produce. "Oh, Johnny is a Novice. This is probably the most I'll get from him," as one teacher put it, is a phrase that is to be avoided at all costs.

The other reason that teachers were leery of making curricular and instructional decisions based on the Kentucky Core Content Test was that they are still skeptical about its value. Not only were teachers skeptical regarding the correlation of scores in science and reading to performance in social studies, but also they were concerned that the scores might not even represent ability in science and reading. Stories of very able students scoring poorly on the test and vice versa have led several teachers to question the value of an individual student's test scores. "I know the students better than the test," was a common response to this line of questioning.

When teachers did report using the test scores of individual students, they typically fell into one of two categories. One group used the scores to identify how students performed on specific released questions. When a large number of students performed poorly on a particular question, the teachers focused on the content of that question in their curriculum. These teachers explained that they hadn't used the scores yet because they had only received an overall score for each student. Many were still awaiting a question-by-question breakdown of the data.

The other tactic teachers employed using the individual student scores was to ability group the students based on the data. This occurred at some schools by classroom, either by grouping students homogeneously by ability in classes, or by purposefully ensuring the heterogeneity of classes by separating similar ability students. At other schools the ability grouping was done in the classroom. Classroom teachers described efforts to make sure that two Novice students were not paired together for peer tutoring or cooperative learning. At the classroom level, none of the participating teachers described a scheme for grouping students homogeneously.

Test Administration Issues

Teacher and Student Testing Requirements

In addition to the challenges represented by the content of the Kentucky Core Content Test, teachers often referred to the administration of the test as a source of undue stress and difficulty. A version of the test was administered to every student in the 4th, 5th, 7th, 8th, 10th, and 11th grades. Students who were absent from school during test administration were required to make up the test when they returned. Only very serious illnesses, verified by a physician, exempted students from taking the test.

This creates a variety of problems for schools and teachers. First, nearly all schools will have some student absences on the days of the test. Those students are required to take the test when they return, so they must miss their normal classes during testing. In addition, they must be supervised during the taking of the test, which occupies staff members beyond the normal testing window.

These requirements do not seem excessive until implemented in a high-stakes accountability system, with severe penalties for test administration errors, that uses the type of test represented by the Kentucky Core Content Test. First, the test is not timed. Students may linger over questions for a very long time. This creates a great deal of added chaos compared with more traditional timed tests. Once the vast majority of students finish, the remaining few must be supervised, and are typically routed to a central location to complete the test. They are not allowed to discuss any part of the test with each other, which requires them to finish each portion of the test in one session. Teachers at one school explained that they were required to bring food in to some students, otherwise completing the test in one session would have caused them to miss lunch. The students ate in isolation from one another, while supervised, and returned to the test immediately afterward.

Second, the test has both a multiple-choice and a writing component. Students may therefore require more materials than the typical number two pencils in order to complete the test. If the students use scrap paper for some form of prewriting or organizing before they commit their answers to their test booklets, that paper must meet with required criteria and be checked by test administrators. Using paper with pre-drawn lines, similar to the paper students use with the 3-4 Column Method of organizing open-response writing, is sufficient to negate the scores of all the students supplied with those papers. These are sheets of paper with two vertical lines drawn on them; no words supplied at all. One of the schools participating in this study was penalized last year for exactly this violation. Ironically, Power Verbs and the 3-4 Column Method are legitimately posted on the walls of nearly every classroom we visited.

The scrap paper, even if it is perfectly within allowable parameters, creates another problem. Test security is difficult to maintain in the best of circumstances. When students construct their answers to a large component of the test outside the test booklet, the problem becomes more severe. Supposedly, the scraps of prewriting are collected and destroyed; however, this procedure is not verified by classroom. If a student has difficulty answering a particular question, it stands to reason that she might bring the topic to the attention of her teachers.

Other problems associated with the administration of the test exist, including the

mechanical problems of getting the test to schools, collecting the booklets once the test is administered, and other organizational and procedural issues. There are also issues that arise sporadically at schools that require special attention even though these instances are isolated and rare. The phone service center (a help line for questions related to testing and accountability) reported that one school described a student becoming so distraught that he became sick and threw up on the test booklet. This is not a situation covered explicitly in the test administration manual. Another school requested permission to collect portfolio materials from a student who was attending the funeral of a close family member. These examples are extreme, but the high-stakes nature of the accountability system has caused some schools to resort to extreme measures to ensure that scores are as high as they can possibly be and that they avoid test administration penalties.

Test Administration for Students with Disabilities

Kentucky was very ambitious in its determination of which students to exclude from taking the state accountability test. It created a standard of inclusion instead of a set of exclusion criteria. Kentucky is one of the very few states that have an alternate assessment system that permits participation by students with severe disabilities. Kentucky's system has been touted as a model for other states in this regard (Ysseldyke, Thurlow, Erikson, Gabrys, Haigh, Trimble, & Gong, 1996). All public school students are to be included in the testing process, including students with disabilities. Those students with learning disabilities so severe that they have no access to the test forms, even with accommodations, are required to complete an alternate portfolio. Those with less severe disabilities are provided large print test booklets, readers, Braille versions of the test, scribes who write for the student, or other accommodations to allow them to complete the test.

In schools with a large population of disabled students, these accommodations require a substantial investment of staff time and often require the school to recruit or hire outside for help. Readers and scribes necessarily work one-on-one with the students throughout a testing process that can span more than a week. Schools are forced to search for the additional staff members required for testing these students.

In addition to the effort required just getting students tested, large populations of students with disabilities gives schools another stigma. The validity and reliability of KIRIS and now the Kentucky Core Content Test have typically been examined on a system-wide scale (Hambleton, Jaeger, Koretz, Linn, Millman, & Phillips, 1995; Koretz & Barron, 1998) or, more recently, for individual students and schools (Hoffman & Wise, 1999). The validity and reliability of the test for students who take it with some form of accommodation is largely unknown (Koretz & Hamilton, 1999).

Test Security

The security of the Kentucky Core Content Test questions is a major concern. Some of the precautions taken to ensure the security of the test include the use of multiple forms, tight control of the testing window, counts of the forms taken before and after testing, secure test form construction, and others. These precautions are necessary because the items can be reused for several test administrations during their life cycle.

The life cycle of a Kentucky Core Content Test item begins once it is written. The procedures for writing and editing test items have varied slightly over the years, but for the sake of argument we will assume that the item is secure at this stage. Before the item can be included in the accountability calculation, however, it must be field-tested. This is done by placing it on an existing form of the Kentucky Core Content Test (and previously on the KIRIS test), but not counting the results of that item as a part of the test. If the item is judged to be of sufficient merit and/or needs only minor editing, it is included as a working item the next year. This means that the item has been "exposed" or sent to schools during testing one time prior to its inclusion. Other items are judged to be of poor quality and eliminated before they can become working items, so exposure of these items is of no consequence. Still other items are judged to be of sufficient merit to be salvaged, although not ready to become a working item. Those items are edited and field-tested again. This means that some items are sent to schools two times before they become working items or are eliminated.

The Kentucky Core Content Test currently has six forms (barring Arts and Humanities and Practical Living/Vocational Studies), with an initial plan to release one form per year[8]. The items on the other forms are recycled and placed on the test again the following year. Therefore, many items will be placed on the test for several years before they are released and retired. A single item could be included on a school's accountability test for as many as five to six years. The degree to which that item can be considered secure/secret is unknown, but certainly suspect (Koretz & Barron, 1998).

Test item security is a major concern for two primary reasons. The first is that the test itself should be a valid and reasonable measure of student ability. This hinges on the test being a fair representation of the skills and knowledge of the students. If the students' teachers were aware of the content of the test, then they might teach toward that content specifically, and omit portions of the curriculum that are not represented by test items. The degree to which this practice is not uniform among teachers and schools creates a bias in the accountability system. Students at those schools more adept at adopting a curriculum that closely matches the live test items would be expected to perform better on the test. The test then becomes a measure of the schools' acuity at mimicking the test items in their curriculum.

Based on comments that teachers volunteered when asked a variety of our questions, we have little doubt that the actual content of the test influences the taught curriculum in many schools. Several teachers commented on specific questions that their students had difficulty with during testing. One elementary teacher told us that she had never taught the phases of the moon to her students specifically. They studied earth science and the moon, but never at the detail level of one of the test questions. She stated that she would definitely cover that topic extensively in the coming year because many of her students couldn't answer the question and it wasn't fair to the students who got that form. Another school's science department complained about the elimination of a specific question. They had purchased equipment to teach an elaborate weather unit because of that question on the test. Since the question was not on the test this year, they felt that the

[8] The release plan has been suspended in order to increase the item pool, meaning even longer life cycles for test items.

equipment and the time spent teaching the topic were wasted. The topic has always been in the Core Content, but the test itself caused the teachers at this middle school to make it a priority in their curriculum.

Communication

Teaching has traditionally been viewed as something that occurs in isolation. A survey conducted in San Francisco public schools in 1972, for example, found that a majority of the principals surveyed reported no day-to-day work relations among teachers of the same grades, and 83 percent of the principals reported no such exchanges among teachers of different grades (Meyer & Rowan, 1983). More recent research continues to call for increased opportunities for teacher interaction, as well. Zemelman, Daniels, and Hyde (1993) state that "…teaching as it is organized in this country has shockingly little room for professional conversation." (p. 168). This isolation can be further reinforced by the lack of external standards against which teachers can measure their students' performance, and by extension, their own schools' performance.

In Kentucky, educators and legislators realized the importance of having an external set of standards when they established a statewide assessment as part of the Kentucky Education Reform Act of 1990. For the first time, all public school students were held to the same standards, and schools were held accountable for their students' performance. As schools struggled to raise their scores to avoid sanctions, it became important to be able to overcome the mindset of isolation. Instead, they needed to communicate with others to find out how best to meet the needs of their students.

In the nearly 10 years since the passage of KERA, some aspects of communication have improved in Kentucky's public schools. For example, some schools in this study have eliminated self-contained classrooms, in which one teacher is responsible for instruction in all subjects. Instead, these schools have created teams of teachers who share responsibility for instructing the same group of students. And some schools now provide their teachers with two planning periods—an individual period and a common period that can be used with other teachers on a team or grade level, or who teach the same subject. Other schools are making an effort to have department-level meetings as well as grade-level meetings. These efforts at improved communication within a school have not translated across schools for the most part, however.

We examined communication problems—and some examples of their resolution—that occurred between schools. They include communication problems between:

- "Feeder" schools within a district, which pass students through a "pipeline" from kindergarten to high school graduation. This might be compared to a factory assembly line in the sense that each level of school must receive a student cohort that has reached a certain academic standard and, in turn, must help that cohort reach standards that prepare it for the next level.
- Same-level schools within a district, such as all the elementary or middle schools within a district.
- Schools and agencies that are outside one's district.

"Feeder" schools

As mentioned previously, accountable-grade teachers must necessarily rely on the teachers in lower grades to prepare their students to meet the rigorous standards exemplified by the Kentucky Core Content Test. Poor communication can impede the smooth transition of students from grade to grade. CATS provides no impetus for schools to support each other's efforts to improve, so when the lower grades are in another school, the problem is exacerbated.

The seventh-grade science teachers at Oak Junior High School offered a solution to this problem. These teachers met this summer with all fifth- and sixth-grade teachers who teach science in the district. The purpose of this meeting was to spell out what students needed to know before coming to the seventh grade, especially in the area of physical science. The seventh-grade science teacher had arranged for a one-day workshop that would review physical science concepts with these teachers and she also prepared a box containing simple physical science experiments and materials for each teacher to use in the classroom. Teachers at Walnut Elementary mentioned similar efforts begun by a high school science teacher. Previously, the high school teacher had visited elementary schools and had worked with their teachers, but the program had to be modified somewhat this year. Instead of visiting schools, the high school teacher presented science workshops to the elementary teachers.

Although teachers in assessed grades complained about the lack of communication, it is safe to say that teachers in non-assessed grades also recognized that communications problems exist. Eighth-grade science teachers in Oak Junior High School, for example, understood that they played no role in the accountability score of their school, since science is assessed at the seventh grade and again at the 11^{th} grade. They also understood that their students were likely to forget what they learned in eighth grade by the time they take the 11^{th} grade assessment. These teachers said they have tried to discuss possible solutions with their high school counterparts, with little success. Instead, they decided to emphasize science skills and procedural methods so students could work more effectively when they got to high school. They recognized the problem and attempted to find a solution for it on their own. However, until the high school science teachers become part of the solution, they will remain skeptical about resolving the larger problem.

Oak's sixth-grade problem will be resolved soon, when a new middle school will replace the old junior high building. The sixth grade will then be housed in the new building and will be considered part of the middle school. Until then, teachers at Oak Elementary are doing what they can to feel connected to the junior high school. A sixth-grade teacher who teaches science and math, for example, spoke of having a mathematics textbook that was aligned with the junior high's curriculum (it was unclear how the textbook selection had taken place), while a fifth-grade science teacher mentioned the workshops that the seventh-grade science teachers presented.

Walnut County's sixth grade problem may not be so easily resolved. Both principals we interviewed in Walnut County recognized the sixth-grade problem during the initial phase of this study, which occurred in spring 1999. The middle school principal said that the sixth grade needed to be at the middle school to help better prepare students for the middle school assessment. However, he said that such a move would be unacceptable to parents of sixth-grade students, who valued having their children stay in smaller, community-based schools rather than busing them to

the middle school, which was perceived as being large and impersonal. The elementary principal had offered to turn over supervision of his sixth-grade teachers to the middle school principal in order to make them feel more connected to the middle school. Nothing had changed during the second phase of the study.

Same-level Schools

While most of the emphasis on school communication appears to be in the "feeder" school arena, we encountered one district in which elementary schools had made efforts to improve communication between them. This was in the Oak Independent district, which has some of the same characteristics of a larger, more urban district—and which must deal with higher levels of student transience than might be expected. (During a 1998 visit to a school in Oak's neighboring county district, a middle school principal stated that many families at her school moved frequently between trailer parks that competed with each other for tenants by offering better deals, such as a month's free rent.)

Moving frequently is hard on students, both academically and emotionally, according to Vail (1996), who reported that one in six third graders has been enrolled in at least three different schools since the first grade. These children who move frequently—regardless of family income—are also more likely to be below grade level in math and English and to repeat a grade, Vail added. Transience has also been shown linked to poor performance on the KIRIS assessment (Medsker, 1998).

But student transience affects the classroom as well as the student. Vail (1996) also noted that teachers with large numbers of transient students find that they get bogged down in trying to "catch up" the transient student to where the rest of the class is, and the instructional pace thus falters. Discipline problems also are more common in high transient classrooms, Vail reported, as some segment of the class always is less familiar with teacher expectations and classroom rules.

The Oak elementary schools are trying to lessen the impact of student transience by working to develop a common elementary school curriculum around four major units. These units would be taught at the same time during the school year. Under this system, those students who move frequently during the school year would be less likely to repeat or to miss a topic when switching schools.

Schools and Other Agencies Outside the District

As mentioned previously, schools had little reason to communicate with agencies or other schools outside their district before the Kentucky Education Reform Act of 1990. Districts operated in isolation, and there were few standards that all had to meet apart from having to spend a certain amount of time on a particular task. With the advent of KERA, however, schools throughout the state were being held to the same standards of performance and accomplishment. Schools and districts were forced to pay attention to rising or falling assessment scores.

We cannot state with assurance that schools are communicating more frequently with other schools and agencies than they previously did. However, we can report that schools in this study

are paying careful attention to their test scores and are using a variety of techniques and resources in the hopes that those scores will improve.

One of the most common outside resources that schools are using is Kentucky's network of regional service centers. These nine centers offer services ranging from assessment score interpretation and analysis to professional development opportunities to operating the Highly Skilled Educator[9] (formerly known as Distinguished Educator) program.

Schools are also using outside programs to guide their instruction and content selection. Programs such as 10 Sigma[10], America's Choice[11], Core Knowledge, and the Accelerated Reading and Math programs all have their adherents who are pinning their hopes on a certain program's ability to raise test scores.

Some schools are encouraging their teachers to visit other, more successful, schools in hopes that they will learn how the other schools succeeded. Cottonwood Middle School social studies teachers did just that—they visited a school that had earned impressive gains on the social studies assessment score. The science teachers were looking forward to a similar visit in hopes of improving their scores, as well. In this example, school administrators were able to recognize the value that such visits can hold, and they were willing to support the visits. Unfortunately, these visits can be expensive, since the school must hire substitutes if their teachers visit during the school year, to watch the strategies in action. While students may be missing some instruction while the regular teacher is gone, the increase in awareness of new approaches and alternative expectations may offset any short-term loss.

Discussion

The discussion section of this report contains reflections, interpretations, and extensions regarding the results of the study. This section is designed to clarify important topics that emerged during the course of conducting the study and analyzing the data. It is important to remember that this study involved only 31 schools from 15 districts, that participation was voluntary, and that the sample was not randomly selected. The degree to which those schools are representative of the system is unknown. The purpose of this section of the report is to provide a basis for discussing important issues in Kentucky schools and to suggest questions that might be asked or even answered during the course of future study.

Teacher Impressions and Reactions

[9] Highly Skilled Educators are provided for schools not making appropriate progress toward improving test scores. They work cooperatively with the school staff to bring Kentucky Core Content Test scores up.

[10] Ten Sigma is both the name of a non-profit company and the generic name of the educational products they produce. They provide a variety of educational materials ranging from lessons, assessments, rubrics, and computer software to training programs for teachers and staff members.

[11] America's Choice was formerly known as the National Alliance for Restructuring Education (NARE). The program is focused on improving education through standards-based education. It is operated by the National Center on Education and the Economy.

The second year of transition between KIRIS and CATS looked much like the first. Teachers remained cautious and skeptical of a new system that supposedly represented an improvement over KIRIS. Most teachers claimed that the new system was better, although not greatly so, while a few remained unconvinced that the system has changed in any significant way at all. A small minority claimed that it was substantially improved.

Given that a large portion of the assessment and accountability system remained unchanged, it is not surprising that large programmatic changes in schools' curricula were not observed. The Kentucky Core Content for Assessment, the curriculum framework on which the assessment is based, remained largely unchanged. Therefore, the content material that students were expected to know did not substantially change. In fact, many of the changes in the system only affected the way each school's accountability index was calculated, not the way students were tested.

Schools had only received data for one administration of the Kentucky Core Content Test at the time of this study and the data they received represents the first year of the interim accountability cycle. Major changes in the ways that schools are run based on that data might not be justified at this early stage in the switch from KIRIS to CATS. Teachers expected more change in the system than they observed, but most did not differentiate between the testing portion of the system and the accountability portion. In a very positive sense, the fact that their practices have not been radically altered may indicate that the curricular and instructional reforms begun by KERA have outlasted the KIRIS testing and accountability system.

Teachers were very positive regarding the reduction in the number of required portfolio entries. Portfolios were typically considered burdensome and time-consuming, and while the reduction in the number of entries may not have reduced the time spent completing portfolios, the burden seems to have been lessened. Teachers were also pleased with the addition of the proficiency sub-categories within the Novice and Apprentice classifications. The extra points for the school and the positive reinforcement for students near cut-points were very well received. However, teachers were very negative regarding the small numbers of students scoring in the Proficient and Distinguished categories. They referred to student percentile scores as evidence of the state's unreasonable expectations. The addition of the multiple-choice items to the school accountability index was typically viewed positively by teachers as well, and many were incorporating more multiple-choice questions into their classroom assessments and everyday assignments. Writing, however, was still emphasized to a great extent. Students spent a good deal of their class time writing and schools expected writing prowess to translate into high portfolio scores as well as high scores on the Kentucky Core Content Test.

Teacher Practices and Perceptions of Student Performance

Wholesale changes in instructional practices were not observed during this study. Teachers were instead fine-tuning, shifting emphasis, altering their order of instruction, and making other small changes in the ways they delivered instruction. A pattern concerning teacher practices was apparent, however. This study, because of its qualitative nature, allows a clarification, or at least an addendum, to a previous HumRRO study (Harris, et al., 1998; Hoffman, et al., 1997). The first time we visited Kentucky schools, we collected a great deal of data from teachers regarding their practices and then compared that data to the school scores on the KIRIS test. We found that

reform-oriented instructional practices were more prevalent in higher-scoring and higher-gaining schools. One interpretation of that relationship was that the practices elicited the higher scores.

Another interpretation became evident during the course of this study. In many instances the practices we observed were not particularly reform oriented. Many teachers stressed memorization and text-skills. Lessons centered on vocabulary, spelling, grammar, and drill-and-practice exercises were common. These lessons were most common in schools and classes containing a large proportion of low performing (as described by their teachers) students. When teachers described programs designed as remediation, they often described practices that were de-emphasized in their own regular classrooms. One school had a remedial math program that was essentially all drill-and-practice. Another had a remedial language arts program that stressed spelling and punctuation. Discussions of these programs with teachers indicated that often, in order to "bring the students up to grade level," teachers used a different approach than that used in non-remedial classes. The traditional practices were considered more effective or more rapid as a means of instructing students who had fallen behind their peers academically. In other words, teachers' perceptions of their students' academic prowess may be affecting their pedagogy.

It is not surprising that teachers would adjust their curriculums depending on the academic skills of their students. It is surprising that they would also adjust their modes of teaching. The result of this pattern is severe. Bradford (1999) describes this form of accommodation as a "pedagogy of poverty." The students who are most in need of strong teaching pedagogy are those taught using the least effective means. Bradford also states in her research that this practice is common in classrooms with a large minority component.

Bradford (1999) lists five standards for effective teaching. They include:

1. Joint productive activity. Teachers and students produce work together cooperatively.
2. Developing language and literacy across the curriculum.
3. Making meaning. Connecting school to students' lives.
4. Teaching complex thinking.
5. Teaching through conversation.

These standards were very evident and common in a few of the high-scoring schools visited during the course of this study. They were rare or absent in many of the lower-performing schools, which raises questions regarding the similarity of students' opportunity to learn depending on what school or class they attend.

This study was not conducted in such a way that correlational analyses of teacher practices and school scores is appropriate; however, this pattern was sufficiently evident that a more definitive research study on this topic would certainly be appropriate. If low performing students are receiving less effective instruction than their peers are the threat to the validity of the accountability system is serious. Expecting schools to post high gains using poor pedagogy is inappropriate. Expecting them to post larger gains than high-performing schools that use more effective teaching pedagogy is even less appropriate.

Unresolved Issues

Teachers' concerns surrounding CATS were much the same as their concerns with KIRIS. Testing different cohorts of students was not generally seen as a fair or equitable way of judging performance. Demographic issues equated to perceptions of an unfair comparison for many schools. Because they serve different populations, broad standards seemed inappropriate to many educators. Test administration was burdensome for schools, even with newly implemented page limits and a testing scheme spread throughout more grade levels.

Accountable Grade Teachers

Teachers in accountable grades were very concerned about being held responsible for students who arrived in their classrooms without the skills necessary to engage in a curriculum designed around the Kentucky Core Content for Assessment. Several teachers expressed that a large portion of their students came to them with below grade-level reading skills and little content area background. The teachers were forced to decide whether to charge forward with their curriculum, leaving many students behind, or to adjust their teaching to the students' current ability level and perhaps not meet the requirements of the assessment system.

There are at least two possible reasons that teachers in accountable grades are more cognizant of these deficiencies. The first is the impending state assessment. These teachers feel so responsible for getting students ready that they are forced to compare their students' abilities with the abilities they feel the students will need in order to score well on the test. It is possible that, although the students' previous teachers evaluated each student prior to their passing a grade level, they might not have done so through so intense a lens. The standards set by Kentucky in each curricular category are high. The state accountability system and the associated student test are reflections of those standards (Hoffman & Tannen, 1998). The rigorousness of those standards becomes much more immediate to teachers closer to the accountability test.

The second reason for the perception of deficiencies in some areas of the curriculum are the measures that some schools have undertaken in order to maximize test scores. Because of weaknesses in non-assessed grades in some subjects, it is very possible that students are indeed coming into classes unprepared. This may happen, especially if the sixth grade is located in the elementary school and the seventh is in the middle school. Elementary students take the science portion of the assessment in fourth grade. They might have experienced an under-emphasized science curriculum in fifth and sixth grades, leaving them in need of much help prior to beginning their seventh grade science curriculum. CATS accountability creates no impetus for the elementary school to help the middle school score well on the test or vice versa. Schools may be working at cross-purposes and making the pressure felt by accountable-grade teachers all the worse.

The seventh-grade science teachers seem to be most "under the gun," but other teachers reported similar stresses. Many seventh-grade teachers reported that students' science instruction in the fifth and sixth grades was either cursory, inadequate, or absent. The assessment was finished by the time the students reached the eighth grade. The inevitable interpretation was that they were solely responsible for preparing the students for the assessment, or that they must cover

the entire three-year curriculum. One school district has reinforced that interpretation by examining scores at the individual teacher level in order to identify instructor strengths and weaknesses.

Portfolios

Portfolios were still seen as overly time-consuming, burdensome, and unreliable for assessment by many teachers. They have a negative stigma that has been reinforced rather than abated by the reduction in portfolio requirements. The elimination of the mathematics portfolio altogether and the reduction in the number of entries in the language arts portfolio has given teachers evidence that the state is willing to reduce the reliance of the accountability system on the portfolio. The willingness of the state to reduce the emphasis of the portfolio was seen as an admission that the portfolio was not valuable or necessary.

The continued emphasis on reading and writing, however, was at least partly due to the continued use of student portfolios as part of the assessment. Schools are responsible for assisting students in creating portfolios and they also assess the portfolios, both during the preparation of each piece in class and in a summative evaluation of all the pieces together as part of the accountability system. Schools therefore felt that they had a great deal of control over what is expected and produced in terms of each student's portfolio work. Their level of influence over portfolio production is considerably greater than their contribution to student test performance. They not only prepared the students to create the portfolios, but they set the standards for what they would accept in class and what had to be redone. It is not surprising that, even though the number of required writing portfolio entries has been reduced and the mathematics portfolio has been eliminated from the accountability index, portfolios still occupied a great deal of time and attention at most schools.

Depth Versus Breadth

Teachers must always decide the appropriate depth of coverage for any curricular topic. The Core Content for Assessment represents a large number of topics and many teachers claim that it forces them to sacrifice depth of coverage for breadth in order to teach them all. Science and social studies teachers were particularly concerned that the scope of the curriculum was too broad to be adequately taught in school (Thacker, Hoffman, & Koger, 1998).

The argument that depth of understanding has been sacrificed for breadth of coverage is not new. It was a major theme of the analysis of the Third International Mathematics and Science Study (TIMSS) data and has been posited as an explanation for the relatively poor showing of United States students compared to other countries (American Association of Colleges for Teacher Education, 1997). The TIMSS data refers only to mathematics and science, but it seems unlikely that the phenomenon is limited to those subjects.

Kentucky teachers have been provided a tremendous amount of information regarding what students are expected to know and be able to do as a result of the reform effort. Teachers were first provided Learner Goals and Academic Expectations (KDE, 1994a) and the two portfolio implementation manuals (KDE, 1994b; KDE 1994c). That information was expanded to include

the Transformations document (KDE, 1995), further clarified in the Core Content for Assessment (KDE, 1996), expanded in the Program of Studies (KDE, 1998), and expanded and elaborated yet again in the Curriculum Implementation Guide (KDE, 1998). Several of these documents have been through revisions and now exist in newer editions. When placed altogether, these documents represent a daunting amount of material. Add the textbook for any given class, supplemental materials provided by textbook manufacturers (workbooks, teachers guides, laboratory manuals, presentation software, videos, etc.), district curricular requirements, any school-wide programmatic curricular material adopted from other companies (Accelerated Reader, Ten Sigma, Core Knowledge, Project Bravo[12], etc.), community expectations, and finally the teachers' own ideas and professional opinions regarding the appropriate curriculum, it is not surprising that the curriculum was described as "too broad." Many middle school teachers were responsible for more than one type of class and many elementary teachers were responsible for teaching the entirety of the curriculum for all subjects. If the teachers served on the planning committees for their schools, they were also responsible for helping create the School Transformation Plan, a required improvement plan for the school, with its own set of regulations and guidelines. The Transformation Planning guidelines have now been replaced by the Consolidated Planning Process (KDE, 1997). The amount of information is so large that it is easy to see how teachers could become overwhelmed.

Another factor that made it all the more difficult to define the Kentucky curriculum was that many of the curricular guides these teachers referred to have been superceded by other more up-to-date versions or altogether different documents. The old documents were not recalled, and indeed teachers often did not know what had become obsolete and what was current. They described their frustration with trying to teach toward a continually moving curricular target. It was perhaps these frustrations that precipitated the call by many of these teachers for a statewide textbook for each assessed class. Most schools seem to have gravitated toward the Core Content for Assessment and use other sources as supplements, but all have been required to choose the particular guide they will emphasize and refer to the others only when a need is perceived.

Communication

Communications within and between schools, as well as communication from outside agencies, is slowly but steadily improving. Curriculum alignment efforts have necessitated that schools work toward system goals in conjunction with school goals. The Kentucky Core Content for Assessment has given many schools a common set of expectations around which to rally. The Core Content has influenced the curriculum of all the participating schools (to admittedly varying degrees). Clearly, schools seem willing to focus their attention on improvement. Test results guide schools in the formation of their Consolidated Plans. Test scores influence professional development, class schedules, programmatic changes, curriculum, and resource allotment.

Curriculum Alignment

[12] Project Bravo! Is a program designed to help students achieve academically and socially. The first portion of the program stresses writing skills while the second consists of lessons that teach the social skills of empathy, impulse control, and anger management.

Curriculum alignment, while providing a guideline for content coverage, may be perceived as a "quick fix" solution to serious communications concerns. When fully implemented, curriculum alignment can provide a "map" that shows each teacher the concept or topics that children are expected to learn in each grade from primary through high school. A properly aligned curriculum ensures that no topics have been accidentally omitted or unnecessarily duplicated, and it also ensures that teachers have a clear understanding of what they and others must teach. In that sense, it should serve as a "contract" that specifies what content students should learn and to what standard. Teachers should be able to count on previous teachers having met their part of the contract, and they, in turn, should assure that their students meet the standards before sending them to the next higher level. Schools in Kentucky commonly align their curriculum with the Kentucky Core Content for Assessment (1996) to ensure that what is eligible for inclusion on the statewide assessment is what teachers are covering in class. A few schools have selected other programs such as 10 Sigma or Core Knowledge in addition to the Core Content. Curriculum alignment is not a substitute for effective communication, but it can serve as a foundation on which to build communication.

Unfortunately, several schools in this study have not used curriculum alignment effectively. While many schools had begun work two or three years ago to align their curriculum, several schools began work only recently, either using it for the first time during the 1999-00 school year or reporting that a draft version was nearing completion during our visit. Still other schools apparently are having problems thinking of curriculum alignment as something that extends beyond their own walls. They mention that they have completed a building alignment, for example, and can even describe communication efforts between colleagues at their school (offering to review the curriculum in the preceding grade level, for example). Yet they do not understand that curriculum alignment must necessarily extend to all levels in the district if it is to be effective. Other schools may have taken the time to complete the curriculum alignment, but have not really implemented it.

Reflecting on our visits to Kentucky schools for the past four years (as part of this study and previous studies related to KIRIS), we sensed a developmental pattern to "curriculum alignment." The first stage, prominent in our initial years, was to align individual classrooms to the Core Content. Because the Core Content, at the time, addressed only student expectations at grades 4, 8, and 11, classrooms in those grades were the ones primarily concerned about curriculum alignment. A second stage emerged that recognized the need to align the content of instruction across grades, but within buildings. Our recent observations suggest that this is the stage many schools are currently in. The third stage, which we began hearing about last year, was for more district involvement in curriculum alignment across buildings. Given the legacy of isolation in school structure (Meyer & Rowan, 1983), and the challenges of KERA (Holland, 1998), this evolution of increasing integration and improving instructional coherence is remarkable. Schools are figuring out that the solution to CATS cannot be pinned on individual teachers, but that a truly system-level solution is required.

The perception by some seems to be that the need for communication has been eliminated if their school has gone through a curriculum alignment process. With the alignment of their curriculum, teachers said that they now knew what they were responsible for covering, as well as what teachers in earlier and later grades were to cover. Yet we heard a "disconnect" between

what they were saying and what they were experiencing when teachers of assessed subjects complained vigorously about their students' lack of preparation in previous grades. Also, teachers frequently spoke of curriculum alignment as a one-time activity to be achieved ("We aligned our curriculum last year."), rather than as an ongoing process of integrating and fine-tuning the curriculum according to current academic standards and student needs.

Arts and Humanities

Several teachers expressed concerns that the wealthier urban schools had a considerable advantage on the Arts and Humanities test. This study was not designed to ascertain the veracity of that statement, but the perception was clearly present in some of the participating rural schools. One rural teacher commented that schools in Louisville and Lexington simply had more access to theater, museums, concerts, etc. Her students had not experienced these things because it was a three-hour bus ride to attend them. She expected scores on the test to be associated more with socioeconomic status and geographic location than with her instructional methodology or prowess. With schools placing continually more emphasis on arts and humanities and practical living/vocational skills, it might be interesting to examine the differences in the relationship between performance and socioeconomic status on those tests compared to other content areas.

Test Administration Penalties

The changeover from the old accountability system under KIRIS to the current interim accountability system under the Kentucky Core Content Test to the newly completed CATS accountability system involved a fairly complex series of mathematical manipulations. It is not surprising that misconceptions regarding the setting of the new baseline have occurred. The added complication of penalties has not been addressed sufficiently to allow the participating principals to adequately prepare their schools for the consequences.

The report from Phase 1 of this study discussed the repercussions of imposing test administration penalties on a school. The demoralizing effect of receiving a negative accountability label and the circumstances surrounding the label cannot help but affect the school staff. Locust Middle School was only beginning to overcome the challenges of dealing with test administration penalties. Last year the penalties were given so much attention that nearly every aspect of the school was affected. Teachers couldn't help but talk about the penalties during interviews, even when the questions were not necessarily related. Teachers spent nearly all of their professional development time in "test ethics" training. They had little faith in the fairness of the system. Many of them had heard stories of much more severe violations that did not result in similar penalties. Still, they went to great lengths to make sure that their teaching and testing procedures were "letter-perfect." If the schools currently expecting penalties respond similarly, the administrations at those schools can expect a great deal of difficulty next year.

Sub-content Scores

Teachers accepted and depended on sub-content scores as instructional tools, typically without understanding how they were generated. This was not surprising given that subject-level scores on the test were generated using an Item Response Theory (IRT) methodology and most

teachers have only a very rudimentary understanding of that process. They instead rely on assurances that the test scores are a reasonable measure of the performance of students and schools. They also attributed the same assurances to the sub-content scores. Despite efforts to ensure that the sub-content scores are as accurate and reliable as possible, they are necessarily less accurate and less reliable than the content scores because they rely on fewer items. The same procedures that are used to maximize reliability on the subject-level tests are inappropriate at the sub-content level. Instead, raw scores (based on the number of correct answers) are used to compute the sub-content scores. This also limits their reliability. Teachers never identified these limits. The cautions regarding the interpretation of the sub-content scores (which are included with the score reports) were clearly not reaching the school level in most cases and not reaching the teacher level in nearly all cases.

Sub-content scores were often over-interpreted and may be causing inappropriate or unjustified programmatic changes at many schools. Schools were altering programs and professional development due to small changes in sub-content scores, despite the warnings and guidelines for interpreting those scores included with the school reports. For example, a school mentioned previously is stressing economics instruction and professional development this year due to a single sub-content score. While this may indeed bring up scores in economics next year, the tradeoffs for implementing the program must affect the rest of the subjects to some unknown degree as well. It affects the other subjects by supplanting class time previously devoted to other topics and teacher professional development time previously spent otherwise as well.

Another school eliminated its culture fair program due to one low sub-content score. In most schools, the "culture fair" would have been heralded as a prime example of reform-oriented instructional pedagogy. It was integrated across the curriculum. Students were actively engaged and presented their findings to one another in a very constructivist manner. The program was engaging for parents and community members and was strongly supported by both. The program had a strong component of promoting tolerance for ethnic and cultural diversity. In this case, however, because of an apparent over-interpretation of a single sub-content score, the program was being replaced with a much more traditional, less reform-friendly, system for addressing the same curricular topic.

There is a requirement, explicit or implicit, for all professional development included in the Consolidated Plan to be "data driven." It has caused schools to be increasingly "data needy." Unfortunately, for reasons of system complexity and competing data sources, "data awareness" has not been particularly reinforced. If we define "data awareness" as an understanding of the origins, value, and appropriate uses of data, many schools exhibit a serious deficit. Several schools purchase diagnostic tests in addition to the Kentucky Core Content Test. A few even purchase entry and exit tests for each core subject. They collect the data from all these tests and incorporate it into their improvement plans. However, when the sub-content score date from the Kentucky Core Content Test and these other measures are contradictory, they look not to the data for explanations, but inward toward the school. One teacher explained that the students had evidently "forgotten" what they knew between taking the diagnostic test at the beginning of the year and the Kentucky Core Content Test in April. The answer was to reorder the curriculum to reinforce the topic judged to be weak in April so that it was covered in March instead of September.

Recommendations Regarding Sub-Content Scores

The over-interpretation of sub-scores is indicative of a problem in the actualization of the testing and accountability system at the school level. Schools, in their efforts to improve, have become more reactionary than the data that they obtain warrants. The result is a mad scramble to patch perceived holes in the various programs of instruction in order to bring up scores. The short-term testing cycle has helped lead schools to focus on short-term solutions to curricular and instructional weaknesses.

A factor that has contributed to this type of thinking was the number of short-term interventions that proved themselves to be effective earlier in the testing program. Tools like the 4-Column Method and Power Verbs[13] were previously described very positively in their ability to raise test scores. These tools focused on improving the organizational writing skills of students, while simultaneously increasing the amount of writing the students did in class. KIRIS, and now CATS, because they have such a large writing component, were very susceptible to these types of interventions. Now students commonly write a great deal in class. Most have already been trained to organize their writing. The same interventions aren't working any more. Schools are being forced to look to other strategies to improve students test scores. Unfortunately, short-term interventions that are effective across all subjects have become much harder to find. So, rather than focus on school-wide shifts in curriculum or pedagogy, schools use the data from the tests they administer to their students to plan interventions by subject or by sub-content in their search for solutions.

This information is not altogether negative. Schools, as they improve, will need specific information in order to fine tune their content area programs to best meet the needs of their students. They seem very willing to accept the state accountability test as their primary source of information for this purpose. By adding a temporal component to the interpretation of the sub-scores and by scaling up the training that surrounds the interpretation of the sub-scores, they could become much more valuable. Instead of schools interpreting their sub-scores on a year-to-year basis, they could be instructed to look for patterns over several years that might indicate a weak curricular area. However, if teachers are going to continue to be provided with the sub-scores, they need to be included in the interpretation training before changes like this have an effect at the school and classroom levels.

The sub-scores themselves also need to be placed under some scrutiny. Estimates of the school-level reliability of the sub-scores should be calculated in addition to the standard error currently reported. The validity of the sub-scores in relation to their uses at the district, school, and classroom level should be explored as well.

Efforts to increase the utility of the sub-scores have been described, including altering the methodology of writing test items. The items are now being designed to target one particular sub-

[13] The 4-Column Method-or a shortened 3-Column version is an organizational tool used for writing answers to open response questions. Power Verbs are a list of verbs and their definitions that describe specifically what task is being asked for by a question. Examples of Power Verbs include explain, compare, contrast, defend, etc.

content strand/topic. This practice should create a more uniform scheme for counting each item toward a sub-score. The current methodology counts items toward either one or more than one sub-score, meaning that some items are much more important than others are for determining those scores.

On the other hand, the test questions designed previously combine more than one strand/topic in a single question, mimicking the integrated curriculum called for by the reform effort. Reducing the scope of the questions is a threat to the intended consequential impact of the testing system. Teachers pay attention to the questions themselves when they plan their curriculum, and a less integrated test may translate into a less integrated curriculum in the future. Put simply, teachers may begin to ask, "Why use an integrated methodology for teaching students when the assessment itself represents discrete pockets of knowledge?" The very existence of the sub-scores promotes the idea that the curriculum is a set of easily distinguishable topics that can be taught or reported on in isolation. The interactions and conceptual overlap of the topics is de-emphasized by reporting the sub-scores and by the redesign of the items. This represents a serious departure from the original goals of the reform effort. An effort should be made to track the changes in curriculum and instruction associated with the sub-scores and the redesign of the test items. Teacher education should also be designed to guard against the erosion of the efforts schools have made to integrate teaching, including discussions concerning the limits of interpreting the Kentucky Core Content Test score reports.

Test Security

Test item security is suspect due to the relatively small number of open-response questions and the relative value of those questions in determining a school's accountability index. Teachers often reported that they were altering their curriculums to incorporate the content of specific test items. Teachers often referred to specific questions when discussing the quality of the Kentucky Core Content Test compared with KIRIS. They frequently described questions that were outside their "taught curriculum." For instance, a fourth-grade science teacher explained that the "phases-of-the-moon question," while implied by the Core Content, was never a part of her particular curriculum. However, since the school scored low on "earth science" she would make sure that it was part of what she taught this year. The long life span of open-response test items makes exposure a threat to the security of the Kentucky Core Content Test and may represent a threat to the overall validity of CATS.

This concern is not trivial (Koretz & Barron, 1998). Teachers are concerned about their students as well as their test scores. When they look at the test, they see the items in terms of whether or not their students are likely to be able to answer them. If they see an item that represents a curricular topic they didn't cover, they rightly assume that their students are unlikely to perform well on that item. Who could blame them for choosing to include that topic in their curriculum next year?

The students themselves further compound the test security issue. When a student does not understand a question in class or on a test, her first inclination is to ask for clarification from the teacher. Even a teacher trying stringently not to break any rules would be hard-pressed not to note the content of the items that his students asked questions about. A teacher might instruct the student

to reread the question and answer to the best of their ability, but to do so without regard for the question at all seems unlikely. Now, if we take this line of thinking to its next logical step, students are more likely to ask questions about items with unfamiliar content. Teachers are likely to adjust their curriculum around the content of the test items themselves. Students are therefore more likely to ask questions about items that are new, that have had no chance to influence the content of the taught curriculum previously. Students, then, draw their teachers' attention toward the items most likely to impact teachers' content coverage. Teachers can't help but notice the new items, and to notice the content of an item is to consider teaching it.

The second reason to maintain strict test item security is that the test represents a sampling of the content implied in the Core Content for Assessment. The test is necessarily smaller in scope than the entirety of the curriculum that students are expected to master during their stint in the public education system. The degree to which teachers teach tested topics to the exclusion of other topics limits the breadth of knowledge that students are exposed to in school. Test scores mask this poverty of curriculum; in fact, it is a response to the content of the test that limits the curriculum.

The new test format limits this problem somewhat by introducing a large number of items into the mix so that it is more difficult for teachers to adjust teaching content to the wide variety of topics included on the test. In fact, there are six forms for each of the core subjects (mathematics, reading, science, and social studies), each with 30 multiple-choice questions and 7 open-response questions, for a grand total of 222 test items. This may seem like too many questions to affect the taught curriculum; however, in reality, only a small number of those questions get very much attention.

First of all, the multiple-choice questions count less on the test than the open-response questions. In fact, it is well publicized that they count only half as much as the open-response items. The "half" is somewhat deceiving as well. There are 24 multiple-choice questions that count on any test form. There are 6 open-response questions. Each multiple-choice item is scored as either "1" for a correct answer, or "0" for an incorrect answer. The open-response items are scored either "0, 1, 2, 3, or 4" depending on how well and how completely the student answers the question. That "raw score" is doubled for the open-response items making each item worth 0-8 points on the raw-score-to-scale-score table. The multiple-choice items are worth from 0-1 point on the same table. The open-response items are therefore worth eight times as much as the multiple-choice items, although the numbers of points on the entire test are split 2/1, open response/multiple choice. In fact, there are only 72 "raw score" points available on any form of the test, so each open-response item counts for one ninth of the total number of points available for each student. Teachers are very much justified in paying attention to the open-response instead of the multiple-choice items.

So, teachers are paying a large amount of attention to only 42 questions. If the teacher is a veteran of the testing process, the number of questions she has already seen reduces that number further. The recycled questions have likely already influenced her curriculum and serve to reinforce those decisions. The number of truly new open-response questions is only 12, or 1 per form where six forms each have A and B versions. These are questions that don't count, but that

will very likely count next year.

This topic deserves considerably more attention. At the very least, a study comparing the apparent difficulty of pretest items during their first and second exposure on the test is necessary. In addition, the test items might be more secure if they were exposed on a rotating shift, so that they don't appear on the test during consecutive years. Clearly, the breadth of the test should mimic the breadth of the curriculum to every extent possible, and that breadth should be expanded by the addition of items. Careful examination of the item pool for content coverage so new items can be targeted toward previously under-represented sections of the Core Content should be coupled with psychometric data analysis. Case study research in representative schools should also provide insight into the extent that test administration is a threat to test item security and the accuracy and validity of the accountability system.

Using Scores for School Planning

Kentucky schools were instructed not to make comparisons between KIRIS and the Kentucky Core Content Test. They were previously instructed to use trend data to create their improvement plans. The interim accountability cycle created a data gap in schools' plans for improving. Many schools relied on continuous input of trend data to judge the effectiveness of various school programs, curricular decisions, and other initiatives the schools might have undertaken. In fact, they were required to document the trend data used to determine the merit of these programs through the Consolidated Planning (1997) process.

This gap in the data has left some serious questions regarding what is to be done with next year's scores. Are schools to begin anew in their search for meaningful trends and only compare next year's data to this year's? Should the schools make comparisons between the trends they had identified before and the current information, or is the current data so different that even those comparisons cannot be made? Will there be any accounting for improvement in setting the new baseline, or will improvements made during the past two years and previously only serve to create even higher expectations for improvement at the school? Comparison with the state mean has left many schools with more questions than answers. Perhaps the most pertinent question is from schools that remained on the same side of the state mean where they started. How different from the state mean should we reasonably be? If we started out 5 points below and are now 7 points below, does that constitute a gain, a loss, or did we maintain our status? This is the question that is most difficult to answer, and it is the one that is most important for schools.

Early Score Release

The clamor to get the scores to schools earlier had led us to believe that there would be more of a sense of urgency about using the scores once they arrived. It may be too early in the change from KIRIS to CATS for teachers and schools to have decided how best to use the scores. The time lag between receiving the KIRIS scores and the possibility of using them to alter school functions is well documented (Thacker, Koger, & Koger, 1998). It may simply be a case of the schools needing some time to adjust to the new system. It may also be a case of the schools needing the summer months to completely digest and react to the test scores. If the former is true, then we can expect the uses of the scores to become more overt and more specific in the next few

years. If the latter is true, the continued rush to get scores to schools may not be justified.

Future Case Studies

Exemplar schools exist where further case study might reveal useful information regarding overcoming obstacles associated with poverty, communication problems, etc. One of the most interesting aspects of this research is that it places the researchers in schools as those schools strive to improve and meet the requirements of a high-stakes testing and accountability system. Invariably, there are schools that participate in this study that have the potential to inform the system. While in-depth school-level case study is beyond the scope of this research, the strategies, successes, failures, programs, etc., could be very informative with regard to the actualization of the Kentucky reform effort.

Bibliography

American Association of Colleges for Teacher Education. (1997). *TIMSS study says U.S. math/science curriculum lacks focus*. [On-line]. Available Internet. http://www.aacte.org/timss.html.

Baldwin, F. (1994). Lighting up Kentucky's fifth. *Appalachia, 27*(3), 4-11.

Borko, H., & Elliot, R. (1999). Hands-on pedagogy versus hands-off accountability: tensions between competing commitments for exemplary mathematics teachers in Kentucky. *Phi Delta Kappan*, 80 (5), 394-400.

Borko, H., Elliot, R., & Uchiyama, K. (1999*). Professional development: a key to Kentucky's reform effort*. (CSE Tech. Report 512). Boulder: UC National Center for Research on Evaluation, Standards and Student Testing.

Bradford, D. J. (1999). Exemplary urban middle school teachers' use of the five standards of effective teaching. *Teaching and Change, 7*(1), 53-78.

Branscome, J. (1970). *Seeking educational opportunity for Appalachian youth.* Paper presented at the West Virginia Council of Colleges on Appalachian Youth, Buckhannon, West Virginia.

Cody, C. B. & Guskey, T. R. (1997). Professional development. In J. C. Lindle, J. Petrosko, & R. Pankratz (Eds.), *1996 Review of research on the Kentucky Education Reform Act*. Frankfort, KY: The Kentucky Institute for Education Research.

Denzin, N. K., & Lincoln, Y. S. (1998). *The handbook of qualitative research*. Thousand Oaks, CA: Sage Publications, Inc.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: an introduction*. (6 ed.). White Plains, NY: Longman Publishers.

Gardner, H. (1983). *Frames of mind: the theory of multiple intelligences*. New York: Basic Books.

Goetz, S. J., & Debertin, D. L. (1997). School finance reform. In J. C. Lindle, J. Petrosko, & R. Pankratz (Eds.), *1996 review of research on the Kentucky Education Reform Act* . Frankfort, KY: Kentucky Institute for Education Research.

Hambleton, R. K., Jaeger, R. M., Koretz, D. M., Linn, R. L., Millman, J., & Phillips, S. E. (1995). *Review of the measurement quality of the Kentucky Instructional Results Information System, 1991-1994*. Frankfort, KY: Office of Educational Accountability, Kentucky General Assembly.

Harris, C. D., Hoffman, R. G., Koger, L. E., & Thacker, A. A. (1998). *The relationship*

between school gains in 7th and 8th grade KIRIS scores and instructional practices in mathematics, science, and social studies for 1997 (HumRRO Report FR-WATSD-98-05). Radcliff, KY: Human Resources Research Organization.

Hoffman, R. G., Harris, C. D., Koger, L. E., & Thacker, A. A. (1997). The relationship between school gains in 8th grade KIRIS scores and instructional practices in mathematics, science, and social studies (LRS97-3). Frankfort, KY: Bureau of Learning Results Services, Kentucky Department of Education.

Hoffman, R. G., & Tannen, M. B. (1998). *Relationships between Kentucky's Open-response Scores for Eighth Grade Students and Their CTBS-5 Scores as Ninth Grade Students*. (HumRRO Report FR-WATSD-98-30). Radcliff, KY: Human Resources Research Organization.

Hoffman, R. G., & Wise, L. L. (1999). *Establishing the Reliability of Student Level Classifications: Analytic Plan and Demonstration.*(HumRRO Report FR-WATSD-99-34). Alexandria, VA: Human Resources Research Organization.

Holland, H. (1998). *Making change: three educators join the battle for better schools*. Portsmouth, NH: Heinemann.

Kelley, C. & Protsik, J. (1996). Risk and reward: Perspectives on the implementation of Kentucky's school-based performance award program. Paper submitted to *Educational Policy*, May 13, 1996.

Kentucky Department of Education, (1999a). *Testing in Kentucky: keys to understanding assessment*. Kentucky Teacher (Special Edition-part1). Frankfort, KY: Author.

Kentucky Department of Education, (1999b). *Testing in Kentucky: keys to understanding accountability*. Kentucky Teacher (Special Edition-part2). Frankfort, KY: Author.

Kentucky Department of Education (1999). *1998-2000 District Assessment Coordinator Implementation Guide for the Commonwealth Accountability Testing System.* Frankfort, KY: Author.

Kentucky Department of Education. (1998). *Program of studies implementation manuals*. Frankfort, KY: Author.

Kentucky Department of Education. (1997). *Kentucky consolidated planning process*. Frankfort, KY: Author.

Kentucky Department of Education. (1996). *Core content for assessment*. Version 1.0. Frankfort, KY: Author.

Kentucky Department of Education. (1995). *Transformations: Kentucky's curriculum framework*. Frankfort, KY: Author.

Kentucky Education Reform Act (Kentucky Revised Statutes 158.645).

Kentucky Department of Education. (1994a). *Kentucky's learning goals and academic expectations*. Frankfort, KY: Author.

Kentucky Department of Education. (1994b). *Kentucky mathematics portfolio teachers' manual*. Frankfort, KY: Author.

Kentucky Department of Education. (1994c). *Kentucky writing portfolio teachers' handbook*. (2nd edition). Frankfort, KY: Author.

Koretz, D. M. & Barron, S. I. (1998) *The validity of gains in scores on the Kentucky Instructional Results Information System (KIRIS)*, MR-1014-EDU. Santa Monica: RAND.

Koretz, D. M., & Barron, S. I., Mitchell, K. J., & Stecher, B. M. (1996). *Perceived effects of the Kentucky Instructional Results Information System (KIRIS)*. Santa Monica, CA: RAND.

Koretz, D. M., & Hamilton, L. (1999). *Assessing students with disabilities in Kentucky: the effects of accommodations, format, and subject*. (CSE Tech. Report 498). Los Angeles: UCLA National Center for Research on Evaluation, Standards and Student Testing.

McDiarmid, G. W., David, J. L., Kannapel, P. J., Corcoran, T. B., & Coe, P. (1997). *Professional development under KERA: Meeting the challenge*. Lexington, KY: The Partnership for Kentucky Schools and The Prichard Committee for Academic Excellence.

Medsker, G. J. (1998). *Determining the Relationship between Student Transience and KIRIS School Results: Are Schools with Transient Students Unfairly Impacted?* (HumRRO Report FR-WATSD-98-12). Radcliff, KY: Human Resources Research Organization.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement (Third Edition)* (pp. 13-104). New York: American Council of Education and Macmillan Publishing Company.

Meyer, J. W. & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology, 83*(2), 340-363.

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. (2 ed.). Newbury Park, CA: Sage Publications, Inc.

Stake, R. L. (1995). *The art of case study research*. Thousand Oaks, CA. Sage Publications, Inc.

Stecher, B. M. & Barron, S. I. (1999). Quadrennial milepost: accountability testing in Kentucky. (CSE Tech. Report 505). Los Angeles: UCLA National Center for Research on Evaluation, Standards and Student Testing. RAND.

Stecher, B. M., Barron, S., Kaganoff, T., & Goodwin, J. (1998). *The effects of standards-based assessment on classroom practices: results of the 1996-97 RAND survey of Kentucky teachers of mathematics and writing*. (CSE Tech. Report 492). Los Angeles: UCLA National Center for Research on Evaluation, Standards and Student Testing.

Thacker, A. A., Hoffman, R. G., and Koger, L. E. (1998). *Course contents and KIRIS: Approaches to instruction and curriculum design in middle school science and social studies* (HumRRO Report FR-WATSD-98-07). Radcliff, KY: Human Resources Research Organization.

Thacker, A. A., Koger, L. E., Hoffman, R. G., & Koger, M. E. (1999). *The transition from KIRIS to CATS: instruction, communication, and perceptions at 20 Kentucky schools* (HumRRO Report FR-WATSD-99-23). Radcliff, KY: Human Resources Research Organization.

Thacker, A. A., Koger, L. E. & Koger, M. E. (1998). *Professional development under the Kentucky Education Reform Act: A study of practices in 30 middle schools* (HumRRO Report FR-WATSD-98-38). Radcliff, KY: Human Resources Research Organization.

Vail, K. (1996). Learning on the move: high student turnover spells trouble for students and schools alike. *The American school board journal, 183*, 20-25.

Wimberley, R. C., & Morris, L. V. (1996). *The reference book on regional well-being: U.S. the black belt, Appalachia* (ED403098): ERIC.

Wolf, S. A. & Borko, H. (1999). "No excuses": school reform efforts in exemplary schools of Kentucky. (CSE Tech. Report 514). Boulder: UC National Center for Research on Evaluation, Standards and Student Testing.

Wolf, S. A., & McIver, M. C. (1999). When progress becomes policy: the paradox of Kentucky state reform for exemplary teachers. *Phi Delta Kappan*, 80 (5), 401-406.

Wolf, S. A., & McIver, M. C. (1998). *Writing whirligigs: the art and assessment of writing in Kentucky state reform.* (CSE Tech. Report 496). Los Angeles: UCLA National Center for Research on Evaluation, Standards and Student Testing.

Ysseldyke, J., Thurlow, M., Erickson, R., Gabrys, R., Haigh, J., Trimble, S., Gong, B. (1996). *A comparison of state assessment systems in Kentucky and Maryland with a focus on the participation of students with disabilities.* State Assessment Series: Maryland/Kentucky Report 1. Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota. (ERIC Document Reproduction Service No. ED 415621)

Zemelman, S., Daniels, H., & Hyde, A. (1993). *Best practice: new standards for teaching and learning in America's schools*. Portsmouth, NH: Heinemann.

Appendix A
Introduction Letters

«Title» «first_name» «last_name»
Principal, «company_name»
«address_1»
«city», «state» «postal_code»

Dear «Title» «last_name»

Thank you for agreeing to take part in the Kentucky Department of Education's four-year program evaluation of CATS, the state's revised assessment program. The cooperation of educators like you is vital to learning all we can about the educational progress being made in our state.

As I mentioned in our recent phone conversation, a two-member research team will visit your school on «date». During the visit, researchers will:

- Interview fourth-, fifth-, and sixth-grade teachers. These interviews will last about 30 minutes, and may be done at your teachers' convenience—before or after school or during a preparation period. We would prefer to interview teachers individually, but we understand how hectic schedules are and we will accommodate small groups of teachers if needed. We also understand that teachers are reluctant to give up their sparse planning time and we are committed to limiting individual interviews to 30 minutes. Group interviews may take a little longer.

- Collect three samples of assessment materials from interviewed teachers—lowest unit of assessment (quiz, daily log assignment, etc.), middle unit of assessment (unit test, chapter test, etc), and the largest unit of assessment they use .

- Interview the principal. This interview could take as long as one hour, but will likely only last about 30 minutes.

- As time permits, observe classroom teaching. In addition to the interviews and the collected material, researchers would like to sit in on some classes. We understand that elementary school may not be divided into clearly delineated classes, but we would like to watch some of the interviewed teachers in action. We will make every effort not to be disruptive and will schedule our observations at the convenience of your teachers. If it is convenient, we would like to see some lessons in science and social studies as part of these observations. However, if scheduling conflicts develop, please give teacher and principal interviews top priority.

We have included several items with this letter to assist you in preparing for our visit.

- Scheduling worksheet. Because our time in your school is limited, we would appreciate your having this worksheet completed before our arrival.

- Letters of introduction for your teachers. Please distribute them at least one week in advance of our visit to give teachers time to collect the requested materials.

- General information sheet, which requests information such as specific directions to your school and recommended motels in your area. We ask that you either mail or fax this sheet to us as soon as possible to help us in planning our visit.

Please don't hesitate to contact me with any concerns or questions. My telephone number is 1-800-219-9030 and my e-mail address is athacker@humrro.org. Again, thank you for agreeing to take part in this important research. We look forward to meeting you and your staff.

Sincerely,

Art Thacker

«Title» «FirstName» «LastName»
Principal, «Company»
«Address1»
«City», «State» «PostalCode»

Dear «Title» «LastName»,

Thank you for agreeing to take part in the Kentucky Department of Education's four-year program evaluation of CATS, the state's revised assessment program. The cooperation of educators like you is vital to learning all we can about the educational progress being made in our state.

As I mentioned to you during our recent phone conversation, a two-member research team will visit your school on (date). During their visit, the researchers will do the following things:

- Interview all science and social studies teachers. These interviews will last about 30 minutes, and may be done at your teachers' convenience—before or after school or during a preparation period. While we prefer to interview teachers individually, we understand that group interviews of teachers of the same subject and same grade may be unavoidable. We realize that teachers are reluctant to give up an entire prep period and we are committed to limiting individual interviews to 30 minutes; however, group interviews will take longer.

- Collect three samples of assessment materials from interviewed teachers—lowest unit of assessment (quiz, daily log assignment, etc.), middle unit of assessment (unit test, chapter test, etc.), and largest unit of assessment (semester test, grading period test, etc.).

- Interview the principal. This is estimated to take about 30 minutes, but could last as long as one hour.

- As time permits, observe science and social studies classes. Because science is assessed in seventh grade and social studies in the eighth grade, we would like to observe a few of those particular classes. However, should scheduling conflicts develop, we prefer that interviews receive priority rather than observations.

We have included several items with this letter to assist you in preparing for our visit.

- Scheduling worksheet. Because our time in your school is limited, we would appreciate your having this worksheet completed before our arrival.

- Letters of introduction for science and social studies teachers. Please distribute them at least one week in advance of our visit to give teachers time to collect requested materials.

- General information sheet, which requests information such as specific directions to your school and recommended motels in your area. We ask that you either mail or fax this sheet to us as soon as possible to help us in planning our visit.

Please don't hesitate to contact me with any concerns or questions. My telephone number is 1-800-219-9030 and my e-mail address is athacker@humrro.org. Again, thank you for agreeing to take part in this important research. We look forward to meeting you and your staff.

Sincerely

Art Thacker

Appendix B
Summaries from Phase 1

THE TRANSITION FROM KIRIS TO CATS: INSTRUCTION, COMMUNICATION, AND PERCEPTIONS AT 20 KENTUCKY SCHOOLS

Summary A
Instructional Practices During the KIRIS-CATS Transition

According to a report by the Human Resources Research Organization (HumRRO), the change in accountability systems from the Kentucky Instructional Results Information System (KIRIS) to the Commonwealth Accountability Testing System (CATS) in the 1998-99 academic year did not cause a great deal of change in instructional practice. HumRRO visited and collected data from 20 schools in 10 districts during the first year of a four-year study of the consequential validity of CATS (Thacker, Koger, Hoffman, & Koger, 1999) before the first administration of the Kentucky Core Content Test. Their findings show that the implementation of CATS has caused a ripple compared to the relative wave of influence the initial implementation of KIRIS caused. Many teachers are convinced that the new accountability system will not be different enough from KIRIS to justify significant changes in their practices. Numerous teachers quipped, "rumor has it that CATS is just KIRIS with a new name." Others contend that there is no way to judge what changes will be justified prior to seeing the test and the first round of results. They are content to "wait and see" at this stage.

The reform movement in Kentucky has generated considerable momentum during the past nine years. Reform-oriented instruction is common. Rote memorization from textbooks has been supplanted by attempts to access higher-order thinking skills. And, with very few exceptions, these changes are viewed positively by teachers. One teacher explained, "Ten years ago there was no guidance and no one to help. Now there are resource persons and performance standards that eliminate the easy way out, that is, always using puzzles and worksheets." Changes in instructional practices and the influence of those changes on KIRIS test scores are well documented (Hoffman, Harris, Koger, & Thacker, 1998; Harris, Hoffman, Koger, & Thacker, 1999). Kentucky's teachers have invested considerable time and effort learning to teach differently since KIRIS began. They are unwilling to abandon those practices without compelling evidence to suggest that they should do so.

CATS includes the Kentucky Core Content Tests, subject-specific tests of student achievement that will be used to incorporate test items in both multiple-choice and open-response formats in each school's accountability index. The KIRIS test also included both types of items, but the multiple-choice items were not used in the school accountability formula. This change in the accountability formula has garnered the most attention in terms of instructional practice, but the changes made with regard to including the multiple-choice questions are minor. Several teachers explained that they had always used multiple-choice questions to some extent and that the addition would not make a difference. Others claimed that they had begun to include more multiple choice items on their classroom assessments, but that their methods of instruction had not changed. One school's principal explained that the Scantron® machine the school had purchased just before the Kentucky Education Reform Act (KERA, the act that began school accountability in Kentucky) passed, was finally seeing some use.

CATS has had a similarly modest impact on teacher professional development. Teachers and individual schools have a considerable amount of choice regarding their professional development (Thacker, Koger, & Koger, 1998), and they are typically not choosing to attend training specific to CATS. The few teachers who did report that they had received any training about the new accountability system had attended either a workshop about test administration ethics or a symposium explaining the proportion for which each part of the CATS accountability system will count. When asked to elaborate on the ethics training, teachers explained, "It's the same training we had for KIRIS." Most were more concerned with the allocation of points used to compute their schools' accountability index. The training they had received often left them with more questions than answers, especially regarding the norm-referenced portion of CATS.

School visits also pointed out several possible factors that may impact further evaluation research concerning CATS. Among those factors are the influence of teacher content knowledge on instructional practices, availability and implementation of reform-friendly teaching materials, self-contained versus departmentalized instruction in the elementary schools, schools' attempts to maximize student content learning in assessment grades, issues related to testing special needs students, and other issues. As CATS becomes more and more a part of the everyday language of schools and as schools strive to earn rewards under the new system, these factors have considerable potential to influence instruction.

Bibliography

Harris, C. D., Hoffman, R. G., Koger, L. E., & Thacker, A. A. (1998). *The relationship between school gains in 7th and 8th grade KIRIS scores and instructional practices in mathematics, science, and social studies for 1997* (HumRRO Report FR-WATSD-98-05). Radcliff, KY: Human Resources Research Organization.

Hoffman, R. G., Harris, C. D., Koger, L. E., & Thacker, A. A. (1997). *The relationship between school gains in 8th grade KIRIS scores and instructional practices in mathematics, science, and social studies* (LRS97-3). Frankfort, KY: Bureau of Learning Results Services, Kentucky Department of Education.

Thacker, A. A., Koger, L. E., & Koger, M. E. (1998). *Professional development under the Kentucky Education Reform Act: A study of practices in 30 middle schools* (HumRRO Report FR-WATSD-98-38). Alexandria, VA: Human Resources Research Organization.

Thacker, A. A., Koger, L. E., Hoffman, R. G., & Koger, M. E. (1999). *The transition from KIRIS to CATS: Instruction, communication, and perceptions at 20 Kentucky schools* (HumRRO Report DFR-WATSD-99-23). Radcliff, KY: Human Resources Research Organization.

Summary B
The Effectiveness of School Communication During the KIRIS-CATS Transition

What had teachers heard about the Commonwealth Accountability Testing System (CATS) prior to administering the Kentucky Core Content Test, a component of CATS? How did they learn about CATS and the Kentucky Core Content Test? These questions helped frame the first phase of a four-year study by the Human Resources Research Organization (HumRRO) concerning the consequential validity of CATS (Thacker, Koger, Hoffman, & Koger, 1999). HumRRO was interested in the possible influences the change from the Kentucky Instructional Results Information System (KIRIS) to CATS might have on instructional practices. In order to evaluate those influences it was important to establish how much the teachers knew about the changes. It was also important to recognize the sources of the information teachers did possess, both to evaluate the effectiveness of established communication channels and to identify the origins of possible misconceptions about the new accountability system.

In the course of visiting 20 schools in 10 districts around Kentucky, HumRRO found that teachers rarely reported that they felt well informed about CATS. When researchers asked teachers what they had heard about CATS, the most common response was, "Not very much." Very few teachers reported having any training concerning CATS. Those who did have professional development meetings about the new accountability system had only attended "test administration ethics training" or a symposium explaining what proportion each component of CATS would count in the computation of the school's accountability index. Those teachers who had attended the ethics training reported that it was identical to the training they had for KIRIS.

Teachers reported that they received their information about CATS from a variety of sources. Most commonly, they learned about the new system from materials placed in their school mailbox, from informal conversations with other teachers, and from the newspaper. Communication between teachers has dramatically improved during the past 10 years. The Kentucky Education Reform Act (KERA) and the associated KIRIS accountability system, stimulated a great deal of teacher interaction, primarily due to curriculum alignment efforts. Those lines of communication, as well as curriculum alignment efforts, are still very much in place.

Teachers often reported that their information came to them in the form of memoranda in their school mailbox. They were often unsure of the origins of these documents, although many assumed that they came from the Kentucky Department of Education (KDE). When we spoke with the principals of the schools, however, we learned that the district office was much more likely to have provided the information. In a way the teachers were not incorrect, because KDE does rely on the District Assessment Coordinators (DACs) as a primary channel of communication to the schools.

Relying on the DACs for getting information to the schools, we discovered, has the potential for two very different problems. The first is that the DAC's time is often very limited in small rural school districts. Schools in these districts typically send representatives to the DAC to collect information for the rest of the school's personnel. Each reinterpretation adds to the possibility that the information will become diluted or altered in some significant way. When schools have specific questions or issues to be addressed in these districts they are often forced to wait for the availability of the DAC. The other end of the spectrum exists in large urban districts. These districts produce so much material for the schools that teachers often do not have the opportunity to sort and interpret the information they receive. Whatever the communication problem, be it a lack of sufficient material or an overabundance of it, the flow of communication from KDE to teachers remains effectively stifled.

KDE maintains a web site on the Internet, complete with e-mail addresses to which teachers and schools can address their questions; however it was rarely if ever mentioned by teachers as a source of information about CATS. The KDE newsletter, the *Kentucky Teacher*, was only very rarely mentioned. Teachers seem to depend on their school to provide them with the information they need. Directly mailing schools may be a more effective choice for distributing important information to teachers than the various other more complex methods currently in place.

Bibliography

Thacker, A. A., Koger, L. E., Hoffman, R. G., & Koger, M. E. (1999). *The transition from KIRIS to CATS: Instruction, communication, and perceptions at 20 Kentucky schools* (HumRRO Report DFR-WATSD-99-23). Radcliff, KY: Human Resources Research Organization.

Summary C
Teacher Perceptions of the Value of the New Testing Program

How confident are teachers that the Commonwealth Accountability Testing System (CATS) represents an improvement over the previous Kentucky Instructional Results Information System (KIRIS)? Researchers from the Human Resources Research Organization (HumRRO) interviewed teachers from 20 schools in 10 districts regarding their confidence in CATS (Thacker, Koger, Hoffman, & Koger, 1999). Although the majority of those teachers could recount the substantive changes to the accountability system, very few stated that they were convinced that the system would be better. A large number of the teachers said that the new accountability system represented a change in name only.

The most common response from teachers was that the new system was still largely unknown. They said that before they could make sound judgments about its worth, they would need to see the test and receive the first round of student scores. The next most common response from teachers can be classified as cautious optimism. The actual responses ranged from "It can't be worse than KIRIS," to "Just adding the multiple-choice questions would make the test less subjective and that would make it better in my opinion."

Another common response from teachers was to relate an opposition to the accountability system on a philosophical level. For these teachers, substituting one reward system with another makes very little difference. They claim that by emphasizing monetary rewards, the system-whether CATS or KIRIS-adds a negative connotation to teaching. "Any time that money is involved someone will find a way to cheat or to play the system," said one of the participating elementary teachers.

Teachers are also not convinced that the change in accountability systems is going to have an effect on the preparation of their students for the next grade level. The overwhelming majority of teachers said "not any" when asked about the influence of CATS on their students. A few teachers said that the addition of multiple-choice questions might give the students some needed practice with the format, which in turn, might help them on other multiple-choice format tests. None professed that their student would know or be able to do more as a result of the change in accountability systems.

Teachers were also asked if they thought that the new system would be fairer than the old one. The majority had very little confidence that the CATS system would help the state administer rewards or assistance more fairly than KIRIS. A few teachers applauded the change in the mechanics of administering rewards from giving rewards as bonuses to administering them as school funds, but they were not convinced that the new system would be better at determining which schools received the rewards in the first place.

House Bill 53 suggests placing students' Kentucky Core Content Test scores (a component of CATS) on their transcripts. Teachers supported this effort toward student-level accountability, but also said they believe that it would make very little difference at the elementary and middle school levels. "It might mean more to high school students, especially if the colleges and universities use the scores to determine admittance or eligibility for scholarships." They were also quick to point out that the students who would be most affected by the addition of scores to their transcripts are the students who are already highly motivated to do well on the test. "The students that don't care about things like grades and transcripts are the ones that need the most attention."

When teachers and principals were asked about their use of the KIRIS score reports from last year, most said that they were used much the same as in previous years. The scores are used as a diagnostic tool for the preparation of specific programs and policies that address each school's improvement goals. Even though KIRIS might be considered a "lame duck" this year, the scores from last year are being examined closely. Schools used those scores to plan professional development and design their Consolidated Plans (KDE, 1997). When principals discussed the scores it became very obvious that the practitioners of public education have not internalized the break between KIRIS and CATS yet. They expect to be able to compare last year's KIRIS scores to this year's Kentucky Core Content Test scores. Many have programs in place that may be bolstered or eliminated because of perceived changes in student scores. The *Kentucky Teacher* (Fishback, 1999) also suggests that attributing growth will be possible using both KIRIS and CATS scores. The words "interim period" were rarely heard during interviews.

Bibliography

Thacker, A. A., Koger, L. E., Hoffman, R. G., & Koger, M. E. (1999). ***The transition from KIRIS to CATS: Instruction, communication, and perceptions at 20 Kentucky schools*** **(HumRRO Report DFR-WATSD-99-23). Radcliff, KY: Human Resources Research Organization.**

Appendix C
Teacher Introduction Letter

Information Sheet for Elementary Teachers
Explanation of HumRRO Research

Background

Since 1997, Kentucky schools have taken part in research studies conducted by the Human Resources Research Organization (HumRRO). The first two studies, which took place in 1997 and 1998, focused on middle schools. The 1997 study investigated how teaching practices were related to KIRIS scores. Researchers found evidence that use of reform practices, such as cooperative learning, extended problem solving, discussion, and student writing, were more likely to increase KIRIS scores than were memorization-based practices. The 1998 study examined how schools use professional development and other teacher preparation activities to adapt to the demands of Kentucky's educational goals.

In spring 1999, HumRRO expanded its research efforts to include elementary as well as middle schools. In the first of a four-year research program, HumRRO researchers visited 10 elementary and 10 middle schools to examine the impact of changing the statewide assessment system from KIRIS to CATS.

For the second round of this four-year effort, we are revisiting the original 20 schools (10 elementary and 10 middle schools) and adding 10 "new" schools (5 elementary and 5 middle schools). Your school has agreed to take part in this study, which will run through 2002. HumRRO researchers will be in your school on Nov. 10. We are asking for your help in a couple of ways:

- We plan to interview all fourth-, fifth-, and (if applicable) sixth-grade teachers at your school, concentrating on science and social studies instruction. Each interview will last about 30 minutes, and can be done at your convenience—before or after school or during a planning period. We would prefer to conduct individual interviews, but we realize how hectic schedules are and we will accommodate small groups of teachers if necessary. We also understand that planning time is in short supply and we are committed to limiting individual interviews to 30 minutes. Group interviews may take a little longer.

- We request that you bring three samples of assessments that we may keep to your interview. These should represent three different levels of assessment: a basic unit of assessment, such as a quiz or log book assignment; a mid-level unit of assessment, such as a chapter test, unit test, etc.; and an upper-level or most comprehensive unit of assessment. Actual samples of student work are not required.

- As time permits, we would like to observe classroom teaching. We understand that elementary school may not be divided into clearly delineated classes, but we would like to watch some of you in action. We will make every effort not to be disruptive and will schedule our observations at your convenience. If possible, we would like to see some lessons in fourth-grade science and fifth-grade social studies.

Confidentiality

We will not identify participating schools or personnel in any report, presentation, or discussion of this research. No information collected by interview, observation, or conversation will be divulged to any administrator, teacher, staff, or student within your school, or to any Kentucky Department of Education staff member. Written reports will provide information in summary form only. However, because your school may have characteristics that make it unique among Kentucky schools, anonymity cannot be guaranteed. If you or other members of the school staff read the report, you may be able to determine that certain parts refer to your school. It is highly unlikely that anyone less familiar with your school would recognize it in the report. Please remember that this research is being conducted to evaluate CATS, not schools.

HumRRO Contacts

Please contact Art Thacker at 1-800-219-9030 (email address: athacker@humrro.org) if you have any questions or concerns.

We look forward to visiting your school and talking with you.

HumRRO research staff
Gene Hoffman
Art Thacker
Lee Koger
Milt Koger

Information Sheet for Science and Social Studies Teachers
Explanation of HumRRO Research

Background

Since 1997, Kentucky middle schools have taken part in research studies conducted by the Human Resources Research Organization (HumRRO). The first study, 1997, investigated how teaching practices were related to KIRIS scores. Researchers found evidence that use of reform practices, such as cooperative learning, extended problem solving, discussion, and student writing, were more likely to increase KIRIS scores than were memorization-based practices. The 1998 study examined how schools use professional development and other teacher preparation activities to adapt to the demands of Kentucky's educational goals.

In spring 1999, HumRRO expanded its research efforts to include elementary as well as middle schools. In the first of a four-year research program, HumRRO researchers visited 10 elementary and 10 middle schools to examine the impact of changing the statewide assessment system from KIRIS to CATS.

For the second round of this four-year effort, we are revisiting the original 20 schools (10 elementary and 10 middle schools) and adding 10 "new" schools (5 elementary and 5 middle schools). Your school has agreed to take part in this study, which will run through 2002. HumRRO researchers will be in your school on Oct. 29. We are asking for your help in a couple of ways:

- Our first priority is to interview all science teachers and social studies teachers at your school. Each interview will last about 30 minutes and will be held at a time convenient for you (before/after school or during your planning period). We prefer to interview teachers individually, but we realize that may not be possible in some situations. If a group interview is necessary, we ask that it consist only of teachers who teach the same subject in the same grade. Group interviews will take longer than 30 minutes.

- We request that you bring three samples of assessments that we may keep to your interview. These should represent three different levels of assessment: a basic unit of assessment, such as a quiz or log book assignment; a mid-level unit of assessment, such as a chapter test, unit test, etc.; and an upper-level unit of assessment, such as a semester test, end-of-grading-period test, etc. Actual samples of student work are not required.

- As time permits, we would like to observe some science and social studies classes. We are especially interested in observing seventh-grade science and eighth-grade social studies, since those subjects are assessed in those grades.

Confidentiality

We will not identify participating schools or personnel in any report, presentation,

or discussion of this research. No information collected by interview, observation, or conversation will be divulged to any administrator, teacher, staff, or student within your school, or to any Kentucky Department of Education staff member. Written reports will provide information in summary form only. However, because your school may have characteristics that make it unique among Kentucky schools, anonymity cannot be guaranteed. If you or other members of the school staff read the report, you may be able to determine that certain parts refer to your school. It is highly unlikely that anyone less familiar with your school would recognize it in the report. Please remember that this research is being conducted to evaluate CATS, not schools.

HumRRO Contacts

Please call Art Thacker at 1-800-219-9030 (email address: athacker@humrro.org) if you have any questions or concerns.

We look forward to visiting your school and talking with you.

HumRRO research staff
Gene Hoffman
Art Thacker
Lee Koger
Milt Koger

Appendix D
Scheduling Worksheets

Elementary Scheduling Worksheet

We have included a worksheet on the following page to help you schedule interviews and observations which will take place during our visit to your school. You will have two researchers available. As we previously mentioned, we would like to interview the following people:

- The principal
- All fourth-, fifth-, and (where applicable) sixth-grade teachers

Please write in the name of the person to be interviewed, along with the time and place of the interview in the appropriate cell. We have found that interviews go more smoothly if they are conducted in a quite place (for example, an empty classroom or conference room, or corner of the media center). Also, please remember that although two interviews can take place at the same time, they should not be scheduled for the same location in order to respect the privacy of the person being interviewed.

Observations will be conducted as time permits. If possible, we would like to observe the following classes:

- Priority to fourth- and fifth-grade classes. If possible, we'd like to observe fourth-grade students doing science and fifth-grade students doing social studies.
- Other observations as available and as time permits

Please keep the following in mind when completing the schedule:

- Both researchers will be available during their one-day visit to your school

- Two interviews can be scheduled for the same time, but not for the same place

- Individual interviews will last about 30 minutes

- If scheduling conflicts arise, please keep in mind:
    1. Principal and fourth- and fifth-grade teacher interviews receive top priority
    2. Fourth-grade science and fifth-grade social studies observations receive next priority

- Some schools may have teaching specialists (science or social studies specialist, for example) who do not fit neatly into a specific grade level or classification. Please use the "other" category on the worksheet if you would like us to interview or observe them.

- Our time in your school is limited. Please help us use that time as efficiently as possible by completing the schedule before our arrival.

Elementary School Scheduling Worksheet. When completing this worksheet, please remember that **two** researchers will be visiting your school. The team can interview two teachers in 30 minutes or observe 2 one-hour classes at the same time. Since they have only one day to complete the visit, they will rely heavily on your help in scheduling these events.

Interviews	1 Principal	1-3 4th Grade Teachers	1-3 5th Grade Teachers	1-3 6th Grade Teachers	Other
Name Time Interview Place					
Name Time Interview Place					
Name Time Interview Place					

Observations	4th Grade Science priority	5th Grade Social studies priority	6th Grade
Name Time Classroom			
Name Time Classroom			
Name Time Classroom			

Middle School Scheduling Worksheet

We have included a worksheet (on the following page) to help you schedule interviews and observations which will take place during our visit to your school. This year's visit will last one day only, and you will have two researchers available. As we previously mentioned, we would like to interview the following people:

- The principal
- All science teachers
- All social studies teachers

Please write in the name of the person to be interviewed, along with the time and place of the interview in the appropriate cell. We have found that interviews go more smoothly if they are conducted in a quite place (for example, an empty classroom or conference room, or corner of the media center). Also, please remember that although two interviews can take place at the same time, they should not be scheduled for the same location in order to respect the privacy of the person being interviewed.

Observations will be conducted as time permits. If possible, we would like to observe the following classes:

- Priority to seventh-grade science and eighth-grade social studies classes
- Other observations as available and as time permits

Please keep the following in mind when completing the worksheet:

- Two researchers will be at your school for one day
- Two interviews can be scheduled for the same time, but not for the same place
- We prefer individual interviews; however, group interviews can be accommodated if teachers teach the same subject at the same grade level
- Individual interviews will last about 30 minutes; group interviews may last a little longer
- Observations will last the entire period to avoid disrupting the class (observations in classes that use block scheduling may be adjusted, however)
- If scheduling conflicts develop, please keep in mind:
    1. Teacher and principal interviews receive top priority
    2. Seventh-grade science and eighth-grade social studies observations receive next priority

- Our time in your school is limited. Please help us use that time as efficiently as possible by completing the schedule before our arrival.

Middle School Scheduling Worksheet. When completing this worksheet, please remember that **two** researchers will be visiting your school. The team can interview two teachers in 30 minutes or observe two one-hour classes at the same time. Since they have only one day to complete the visit, they will rely heavily on your help in scheduling these events.

Interviews

	1 principal	**1-3 6th science teachers**	**1-3 7th science teachers**	**1-3 8th science teachers**	**1-3 6th social studies teachers**	**1-3 7th social studies teachers**	**1-3 8th social studies teachers**
Name Time Interview place							
Name Time Interview place							
Name Time Interview place							

Observations (If possible, please give priority to 7th-grade science and 8th-grade social studies.)

	6th Science	**7th Science Priority**	**8th Science**	**6th Social Studies**	**7th Social Studies**	**8th Social Studies Priority**
Name Time Classroom						
Name Time Classroom						
Name Time Classroom						

Appendix E
Logistics Information

Please fax or mail the following information:

TO: Art Thacker

FAX: (270) 351-3620

MAIL: 295 W. Lincoln Trail Blvd.
Radcliff, KY 40160

FROM:

1. In addition to the principal, who else may we contact, particularly if we need to make last-minute adjustments for bad weather? Are there alternate phone numbers available?

Name	Alternate Phone Number

2. Please give us directions to your school or sketch a map:

3. Do we need to know anything about parking restrictions at your school?

4. Can you recommend a motel near your school? (Name, location, phone number)

5. What time does your school day begin (what time do we need to arrive)?

Appendix F
Teacher Interview Topics

Teacher Interview Questions (Fall 1999)

Hello, my name is _____. I am a researcher with Human Resources Research Organization (HumRRO). HumRRO is a private, non-profit research organization under contract to the Kentucky Department of Education (KDE).

HumRRO's Task

Our task is to collect and analyze data on validity issues associated with the Commonwealth Accountability Testing System (CATS) and Kentucky Core Content Test (KCCT). We are independent from the test developers and report directly to the Kentucky Department of Education (KDE) with our findings. We have worked with KDE investigating validity issues with KIRIS for the past three years. This is the second of a four-year project investigating the validity of the new tests.

Thanks (Begin again here)

Thank you for taking the time to complete this teacher interview.

Objectives

We want to hear

- what you think about the new accountability and testing system and
- how you believe it will affect you and your students.

We also want to get a better understanding of

- your instructional approaches and
- activities you use in your classroom.

Perceptions

I will be asking for your perceptions about CATS.

Confidentiality

- We will not identify participating schools or personnel in any report, presentation, or discussion of this research.
- However, because your school may have characteristics that make it unique among Kentucky schools, anonymity cannot be guaranteed. If you or other members of the school staff read the report, you may be able to determine that certain parts appear to refer to your school. It is highly unlikely that anyone less familiar with your school would recognize it in the report.

Target of Evaluation

Please remember that this research is being conducted to evaluate CATS, not teachers, schools, or districts.

Evaluation

Length of Interview The interview was designed to take less than 30 minutes.

Do you have any questions before we begin?

The first section is designed for us to collect some background information on you.

- ***Note: Researcher will read back what you have recorded for verification.***

1. What subject and/or grade(s) do you teach this year?	
Subject	Grade

2. How long have you taught this (these) grade(s) and/or subject(s)?	
Grade(s)	Subject(s)

Has there been any reorganization or major change at the school in the last two years? (Such as some grades moving to another school, changing to team teaching, block scheduling, redistricting, getting a new administration at the school, or major staff changes.)

3. Counting this year, how long have you been teaching at this school?	

4. Counting this year, how many years have you been teaching in Kentucky? Elsewhere?	In Kentucky	Elsewhere

1. How do you think your school performed on the (KCCT) test? What about your content area specifically?

2. Are any content areas within your grade level currently receiving special attention or emphasis? If so, could you discuss why you think the school chose to target those areas instead of others?

3. Are you participating in the implementation, continuation, or elimination of any specific programs within your content area(s) as a response to test scores? Please describe those programs. *(Reword as necessary based on the answers to 1 and 2)*

4. Now that the school's given the test and received the scores, do you plan to change your instructional practices because of CATS? (Are there practices or content that will be emphasized more/less now?)

5. Is the school altering or creating any professional development opportunities in response to the test scores?

6. Has the reduction in portfolio requirements had an impact in your classroom?

7. Does the school have any programs targeting students in the lower (Novice and Apprentice) academic categories? Are those programs specific to any particular content areas?

8. For 5th and 8th Grade Teachers only: Have you seen the individual student scores for your current students? Do you use the students' scores from 4th/7th grade for classroom instructional decisions? Describe.

9. The Kentucky Core Content for Assessment has been revised recently. Have you had a chance to look it over? *(ask the rest only if "yes")* Have the changes had any impact on your classroom practices? In your opinion, do the changes to the Core Content make it better, worse, or is it about the same?

10. Has receiving the test scores earlier improved their usefulness? How so?

11. Do you plan to do anything else differently as a result of test information or do you have comments or concerns that we haven't already discussed?

Appendix G
Principal Interview Topics

Principal Interview

1. How do you think your school performed on the test? Did student performance meet your expectations? Were you surprised by any of the results?

2. Tell me about this year's preparation for CATS. What specific things are being/have been done at the school to get teachers and students ready? Tell me how each of these programs got started.

3. Is any content area or grade level receiving special attention this year? Why is it receiving special attention? Describe what is being done.

4. Is that any different from last year's preparation for CATS? How?

5. How are the initial CATS score reports being used this year? Are the faculty and administration using the score reports differently from KIRIS? Describe. *(Ask only if this question has not been fully explored during the previous 4.)*

6. Have you seen or is the school promoting any changes in teaching practices associated with switching to CATS? Describe them.

7. Are there changes in instructional practices that you would like to see occur at the school? Describe.

8. Does the school have any program specifically designed for lower-scoring students (Novice and Apprentice)? Describe.

9. I'd like to ask you about teachers attitudes toward testing under CATS. (a) Have there been more/fewer/about the same complaints about testing in general? (b) About fairness? Have any new concerns regarding testing/fairness emerged with the change in testing systems?

Appendix H
District Assessment Coordinator Interview Topics

My name is _____. I'm with the Human Resources Research Organization, or HumRRO. As you may know, our company is doing some research on the CATS test for the Kentucky Department of Education, and some schools in your district have been helping us out with that research.

- This is taking place over a four-year period, and last spring was the first year of the project. We visited 10 districts around Kentucky, visiting an elementary and a middle school from each of the 10 districts. We also talked to district personnel. We asked about their perceptions about the transition from the old KIRIS system to the CATS system.
- This fall we made the second round of visits to the 10 districts, and we added 5 new districts—we wanted to time our arrival while the CATS scores were still fresh in everyone's mind.
- There will be two more rounds of visits, one in spring 2001 and one sometime after that.
- All responses are confidential—we don't identify districts, schools, or individuals by name.
- The purpose of our research is to evaluate the CATS test, not individuals, schools, or districts.

We'd like to get some input from district level people for this year's study, similar to what we did in the spring. Would you have time now to answer a few questions about the CATS test?

One of the questions that we asked principals concerned the role that the school played in helping students and teachers prepare for the coming CATS test. I'd like to ask the same basic question—from the district's point of view. So here's the first question…

1. Describe the district's role in helping your schools prepare for the coming CATS tests—what specific things are being/have been done at the district level to help get the teachers and students ready? Does that differ from what the district did for schools during KIRIS?

2. What role, if any, does the district play in the CATS score analysis?

3. What aspects of CATS do you view as strengths? As weaknesses? Do these strengths and weaknesses make the Kentucky Core Content Test a better or worst test than KIRIS was? Why?